Corruption and calcified hierarchies: How Institutional Quality shapes intergenerational mobility in Italy

Immacolata Prisco

Dissertation written under the supervision of Joana Silva

Dissertation submitted in partial fulfilment of requirements for the MSc in MSc in Economics – Major in Macroeconomic Policy, at Universidade Católica Portuguesa and for the MSc in Economics at Università degli Studi di Napoli Federico II, June 2024.

Resumo

Estimamos a mobilidade intergeracional na Itália e examinamos o papel da corrupção nos padrões de mobilidade observados. Utilizando dados cross-sectional do Survey on Household Income and Wealth (SHIW) do Banco de Itália (2004-2022) estabelecemos estimativas baseline de mobilidade nas dimensões educacional, ocupacional e de rendimento. Empregamos o método Two-Sample Two-Stage Least Squares (TSTSLS) para imputar o rendimento dos pais e a metodologia Two-Stage Least Squares (2SLS), instrumentalizando a corrupção com indicadores regionais de cheating e between-class test score variability, para identificar o impacto causal da corrupção na mobilidade educacional. Para investigar os mecanismos subjacentes, utilizamos modelos multinomial logistic para testar os efeitos da corrupção na occupational inheritability e na education devaluation. Os resultados mostram persistência intergeracional substancial: a nossa estimativa corrigida de income elasticity indica que 38% das vantagens de rendimento são transmitidas entre gerações, enquanto o rank-rank slope de 0,362 implica que filhos de famílias do décimo superior classificam-se 29 percentis acima daqueles do décimo inferior. A corrupção fortalece significativamente a transmissão intergeracional da educação. Um aumento de um standard deviation na corrupção reduz o nível educacional em 1,02 anos para filhos de famílias sem educação, enquanto cada ano adicional de educação parental compensa este efeito em 0,056 anos. A corrupção aumenta a occupational inheritability, protegendo o estatuto de elite e desvalorizando a educação como ferramenta de mobilidade.

Autora: Immacolata Prisco

Título: Corrupção e hierarquias calcificadas: Como a Qualidade Institucional molda a mobilidade intergeracional na Itália

Palavras-chave: Mobilidade Intergeracional, Corrupção, Desvalorização da Educação

Abstract

We estimate intergenerational mobility in Italy and examine the role of corruption in observed mobility patterns. Using cross-sectional data from the Bank of Italy's Survey on Household Income and Wealth (SHIW) spanning 2004-2022, we first establish baseline mobility estimates across educational, occupational, and income dimensions. Given the lack of parental income information we employ Two-Sample Two-Stage Least Squares (TSTSLS) to impute fathers' income from retrospective characteristics reported by sons. Then, to identify corruption's causal impact on educational mobility, we employ Two-Stage Least Squares (2SLS) methodology, instrumenting corruption with regional cheating indicators and between-class test score variability. To investigate the underlying transmission mechanisms, we use multinomial logistic models through which corruption's effects on occupational inheritability and education devaluation are tested. Results show substantial intergenerational persistence: our corrected income elasticity estimate indicates that 38% of income advantages are transmitted across generations, while the rank-rank slope of 0.362 implies children from top-decile families rank 29 percentiles higher than those from bottom-decile families. Examining corruption's role, we find it strengthens the intergenerational transmission of education. A one standard deviation increase in corruption reduces educational attainment by 1.02 years for children from uneducated families, while each additional year of parental education offsets this effect by 0.056 years. Investigating the underlying transmission mechanisms, we find corruption increases occupational inheritability, primarily protecting elite status and devalues education as a mobility tool.

Author: Immacolata Prisco

Title: Corruption and calcified hierarchies: How Institutional Quality shapes intergenerational mobility in Italy

Keywords: Intergenerational Mobility, Corruption, Education Devaluation

Contents

1. Introduction

Intergenerational mobility represents the relationship between individuals' socioeconomic status and that of their parents, serving as a metric for evaluating equality of opportunity. High mobility indicates a system where achievement depends primarily on personal effort rather than inherited advantages, making it a cornerstone of meritocratic ideals (Corak, 2013; Chetty et al., 2014). When mobility is low, socioeconomic advantages and disadvantages become entrenched across generations, undermining economic efficiency and social cohesion.

Italy exhibits substantially low intergenerational mobility across multiple dimensions. Recent evidence using administrative tax data reveals limited income mobility: the rank-rank slope of 0.22 (Acciari et al., 2022) means children from families at the 90th percentile of the income distribution rank approximately 18 percentiles higher than those from the 10th percentile families. Educational mobility is similarly constrained, with Van der Weide et al. (2024) positioning Italy among countries with moderate to lower educational mobility compared to other high-income nations. Italy represents an ideal laboratory for studying the specific relationship between corruption and mobility due to its substantial regional variation in both corruption levels and mobility outcomes (Acciari et al., 2022).

This thesis provides the first comprehensive **analysis of intergenerational mobility in Italy across educational, occupational, and income dimensions** and assesses **corruption's role in observed mobility patterns**. For educational mobility, we employ the same indicators used by Van der Weide et al. (2024), which capture different aspects of mobility - from short-range movements between adjacent education levels to long-range transitions from the bottom to top of the distribution. For occupational mobility, we follow Long and Ferrie (2013), who used the Altham statistic, which shows how much occupational inheritance deviates from what we would expect if jobs were randomly assigned, thus capturing true intergenerational association beyond structural economic regional differences. Before quantifying income mobility, we implement the Two-Sample Two-Stage Least Squares (TSTSLS) methodology to impute fathers' income from their reported characteristics, since survey data lacks direct information on parents' earnings.

To investigate corruption's causal impact on educational mobility, we employ Two-Stage Least Squares (2SLS) estimation, instrumenting regional corruption levels with novel measures of rule-bending behavior: standardized test cheating indicators and between-class score variability that capture cultural norms enabling corrupt practices. For corruption’s effect on occupational mobility, we employ multinomial logistic models that allow corruption to differentially affect transitions across multiple occupational categories, revealing how it shapes career pathways. Beyond documenting corruption's effects on mobility outcomes, we investigate the underlying transmission mechanisms through which corruption operates: enhanced occupational inheritability and education devaluation.

Our analysis reveals substantial intergenerational persistence across all dimensions. Educational mobility indicators show significant regional variation, with Southern regions consistently exhibiting stronger persistence than Northern areas. Occupational mobility patterns demonstrate qualitatively different mobility regimes across macro-regions: while Northern Italy's barriers center on educational-entrepreneurial segregation (teachers' children are 80 times more likely to remain in education than become entrepreneurs), Southern Italy exhibits closed entrepreneurial networks where blue-collar workers' children are 200 times more likely to remain manual workers than become entrepreneurs. Our income mobility estimates yield an intergenerational elasticity of 0.376, indicating that approximately 38% of income advantages transmit across generations, while the rank-rank slope of 0.362 implies children from top-decile families rank 29 percentiles higher than those from bottom-decile families.

Examining corruption's role in these patterns, we find it strengthens intergenerational transmission. A one standard deviation increase in corruption reduces educational attainment by 1.02 years for children from uneducated families, while each additional year of parental education offsets this effect by 0.056 years, effectively protecting elite children while harming disadvantaged ones. Corruption enhances occupational inheritance, reducing elite children's probability of downward mobility by 79% while cutting manual workers' children's chances of accessing managerial positions by 36%. Investigation of transmission mechanisms reveals corruption operates through two complementary mechanisms. First, corruption enhances the direct "inheritability" of occupational status: when professional positions become inherited, educational trajectories naturally align with anticipated occupational outcomes. Second, corruption devalues education as a mobility instrument by weakening the link between educational credentials and labor market success - less-educated families rationally reduce educational investments when observing diminished returns, while highly educated families maintain strong educational aspirations regardless of local conditions.

This work contributes to several literatures. First, it extends the **educational mobility literature** (Van der Weide et al., 2024; Neidhöfer et al., 2018; Hertz et al., 2007) by moving beyond Van der Weide et al.'s (2024) cross-country correlational evidence linking institutional quality to educational mobility. We provide the first causal identification of corruption's specific impact on educational mobility patterns, leveraging corruption's role as a key institutional quality dimension. Second, it contributes to the **corruption measurement literature** (Nifo and Vecchione, 2014; Golden and Picci, 2005; Tabellini, 2010) by introducing novel behavioral instruments that capture cultural norms enabling corrupt practices. Our cheating indicators and between-class test score variability provide new ways to measure corruption's cultural foundations. Third, it advances the **TSTSLS methodology literature** (Piraino, 2007; Mocetti, 2007; Orihuela et al., 2023; Solon, 1992) by applying bias correction techniques that yield the most accurate SHIW-based estimates to date. This enhanced accuracy is evidenced by our results' convergence with the Acciari et al. benchmark derived from administrative data.

Beyond these methodological innovations, our empirical findings extend existing knowledge in two areas. We contributes to the **occupational mobility literature** (Long and Ferrie, 2013; Erikson and Goldthorpe, 1992) by linking granular occupational mobility patterns to specific institutional determinants rather than just historical trends. We are the first to apply multinomial logistic modeling in this context, innovatively disentangling corruption's differential effects across specific occupational transitions. Finally, we extend the **education devaluation literature** (Mocetti and Orlando, 2017; Lindner, 2014; Borcan et al., 2017) by building on Orlando's findings that corruption makes education less predictive of labor market success in the public sector. We broaden this result by testing the devaluation across more disaggregated occupational categories and emphasising the resulting consequences for mobility.

This thesis is organized as follows. Section 2 introduces the data sources and sample selection. Section 3 examines educational mobility in Italy, beginning with descriptive patterns and then analyzing corruption's effects on educational persistence through 2SLS estimation. Section 4 investigates occupational mobility patterns in Italy using Altham statistics and develops multinomial logistic models to test occupational inheritance and education devaluation in corrupt environments. Section 5 estimates income mobility using bias-corrected TSTSLS methodology. Section 6 concludes.

2. Data Sources and Sample Selection

This study employs data from the Survey on Household Income and Wealth (**SHIW**) conducted by the Bank of Italy, focusing on eight waves spanning from 2004 to 2022. Previous studies have used SHIW data to estimate mobility in Italy (Piraino, 2007; Mocetti, 2007), focusing exclusively on income transmission. The SHIW provides detailed information on socio-demographic characteristics of Italian households and, crucially for our research, retrospective information about respondents' parents, with the notable exception of parental income. Our analysis exclusively utilizes data from household heads, as they are the only respondents for whom retrospective parental information is collected. For households appearing in multiple survey waves, we retain the most recent observation to avoid double-counting, unless there was a change in the household head during the survey period.

We implement distinct sample selection criteria tailored to each type of mobility measure.
For **educational mobility** analysis, we include individuals aged 21 and above to ensure most respondents have completed or are nearing completion of their educational path. This criterion results in the elimination of very few observations, as individuals younger than 21 are rarely household heads. Current university students are assigned tertiary education status, reflecting the high probability of degree completion for individuals who have persisted in higher education beyond age 21. This helps avoid underestimating educational attainment for younger cohorts still engaged in their studies, following Van der Weide et al. (2021). The educational mobility analysis uses the highest educational attainment between parents as a proxy for the family's human capital endowment. The **occupational mobility** sample includes individuals aged 30-55, a standard age range in mobility studies designed to minimize life-cycle bias. Respondents report both parents' occupations and, to maintain methodological consistency with our educational mobility analysis, we employ the occupation of the parent with the highest educational attainment for our occupational mobility calculations. Samples descriptive statistics are shown in Table A1 Panel A of the appendix. **Income mobility** analysis requires additional data from earlier SHIW waves (1989, 1991, 1993) to construct the auxiliary sample of pseudo-fathers necessary for our Two-Sample Two-Stage Least Squares (TSTSLS) methodology. Our focus on father-son pairs, rather than including mothers and daughters, is motivated by the persistent gender gaps in labor market participation and earnings in Italy, potentially confounding our intergenerational mobility estimates. For pseudo-fathers, we restrict the

sample to male family heads aged 35-50 with positive income, as zero incomes poorly proxy the permanent income measure ideally required for mobility calculations. For sons, from the more recent waves, we use only observations with complete retrospective information about parents. The sample is limited to employed male household heads aged 35-50 with positive income. To maximize methodological validity across both datasets, we standardize key variables: occupational status, economic sectors and educational attainment. Crucially, we restrict the sons' sample to ensure their reported fathers' birth years align with those recorded in the pseudo-fathers' auxiliary sample, maintaining the cohort overlap necessary for valid TSTSLS estimation.

We employ the "**Control and Corruption**" sub-index from the Institutional Quality Index (**IQI**) built by Nifo and Vecchione. This index combines three objective indicators: (i) crimes against public administration, (ii) the Golden-Picci discrepancy between physically existing public infrastructure and allocated funds and (iii) counts of municipal administrations placed under external administration due to inefficiency or corruption. Available at regional level for 2004-2019, we invert the normalized scale for intuitive interpretation, where higher values represent greater corruption: Corruption = 1 - Control and Corruption Index. Our corruption-mobility analysis focuses on the most recent cohort (born 1980-2001) whose critical educational decision-making period (approximately age 16) occurred between 1996 and 2017, ensuring substantial temporal overlap with our corruption data.

To isolate the effect of corruption on mobility outcomes, we incorporate some regional controls that capture underlying socioeconomic conditions that might independently affect intergenerational mobility. Our selection of control variables follows Acciari et al. (2019), who identified key regional characteristics strongly correlated with absolute mobility patterns across Italy (descriptive statistics are presented in Table A1 Panel B in the appendix):

Employment Rate: Measures the percentage of individuals aged 20-64 who are employed relative to the total population in the same age group. This indicator captures labor market dynamism and opportunity structures at the regional level. Data is provided by ANAC (National Anti-Corruption Authority) from official labor market statistics.

Income Level: Calculated as the ratio between total gross family income and family size (in euros). This measure reflects regional economic prosperity and living standards, which directly influence investment capacity in human capital. This indicator is sourced from ANAC.

Income Inequality: Measured using the Gini coefficient for dependent labor income, ranging from 0 (perfect equality) to 1 (maximum inequality). Higher inequality has been linked to lower intergenerational mobility in the "Great Gatsby Curve" literature. Data is provided by the Ministry of Economy and Finance and normalized by ANAC for regional comparative analysis.

Separation Rate: Measured as separations per 100 marriages and sourced from ISTAT. This serves as a proxy for family instability that may influence both human capital formation and the effectiveness of intergenerational transmission mechanisms. Interestingly, separation rates tend to be higher in regions with otherwise favorable socioeconomic conditions.

Educational Quality: Extracted from the IQI index, this indicator is based on average INVALSI test scores (Italian, Mathematics, and English) at primary and secondary school levels. The index captures the quality of educational services provided at the regional level, which affects human capital formation and educational mobility patterns.

To address potential endogeneity in our corruption measure, we employ two instrumental variables that capture cultural norms related to rule-following, but are unlikely to directly affect mobility outcomes except through their correlation with corruption:

Cheating Indicator: A regional-level measure of standardized test result inflation (0-1 scale), constructed using four parameters: abnormally high correct answer percentages, low within-class response variability, suspicious response pattern homogeneity, and low missing response rates. This indicator captures the regional propensity toward rule-bending behavior.

Between-Class Variability Indicator: The percentage of schools where test score variance between classes exceeds the national average (Grade V Mathematics INVALSI Test). This indicator reflects violations of the "equi-heterogeneity" principle in class formation, where high between-class variance suggests non-random student assignment potentially reflecting patron-client relationships characteristic of corrupt institutional environments.

To incorporate these multidimensional regional controls while avoiding multicollinearity issues, we implement Principal Component Analysis (PCA). Prior to this analysis, we transform the income inequality measure (Gini coefficient) using its inverse to ensure consistent directional interpretation across all indicators, with higher values uniformly representing more favorable socioeconomic conditions (remarkably, the separation rate is positively correlated with economic prosperity). The first principal component captures approximately 88% of the total variance in these regional indicators, providing a parsimonious yet comprehensive measure of regional socioeconomic conditions. This composite regional control variable is included in all regressions, enabling us to distinguish corruption's effects from those of broader regional development disparities.

3. Educational Mobility in Italy

3.1 Patterns Across Regions and Cohorts: Descriptive Analysis

Graphs in Figure 1 present a multidimensional analysis of educational mobility in Italy, following van der Weide et al. (2024).

Correlation Mobility (1-COR) equals one minus the correlation coefficient between parent's and children's years of schooling, with higher values indicating weaker intergenerational persistence. The graph shows the South consistently exhibiting lower values (around 0.40-0.45), reflecting stronger educational persistence in Southern regions. The North shows an improvement from the earliest cohort (0.35) to the most recent (0.50).

Regression Mobility (1-BETA) equals one minus the regression coefficient from a linear regression of children's years of schooling on parents' years of schooling. The graph reveals steady improvement across all regions from low initial values (0.17-0.30 for the 1905-1929 cohort) until the educational expansion cohort (1960-1969), followed by convergence for the most recent cohort (1980-2003). While Northern and Central regions reach values around 0.55-0.57, the South stabilizes at a slightly lower level of approximately 0.50.

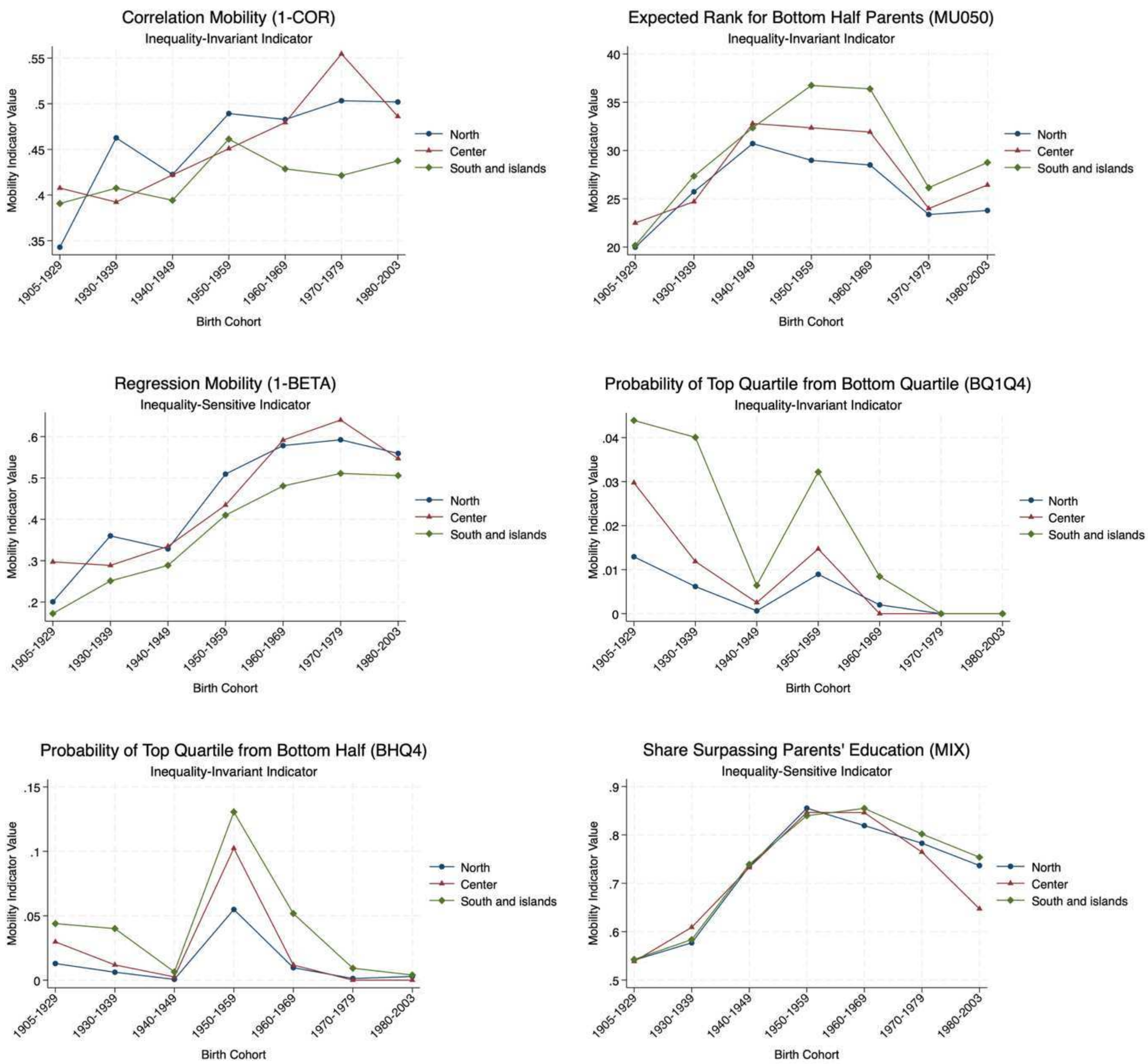
Share Surpassing Parents' Education (MIX) captures the proportion of respondents achieving higher education than their parents or who reached tertiary education like their parents, accounting for ceiling effects where highly educated parents' children cannot show upward mobility beyond a certain point. The inverted U-shaped pattern peaks during the economic boom (1950-1959) around 0.85, with Southern regions showing a "Southern mobility paradox" reflecting greater room for advancement from lower baseline education levels.

Expected Rank for Bottom Half Parents (MU050) represents the expected percentile rank of children whose parents are in the bottom half of the distribution. Under perfect mobility, it would be 50. The graph shows Southern regions' higher values during 1950-1969, suggesting enhanced opportunities for disadvantaged families during public education expansion.

Probability of Top Quartile from Bottom Half (BHQ4) and Probability of Top Quartile from Bottom Quartile (Q1Q4) show complementary patterns of long-range educational mobility. BHQ4 measures children from the bottom half reaching the top quartile, peaking during the 1950-1959 cohort (South: 0.13, Center: 0.10, North: 0.05) before converging to near zero for recent cohorts. Q1Q4, which, absent in van der Weide et al., we introduce to complement BHQ4, captures the more extreme mobility from the bottom quartile to the top quartile. It shows initially higher values in the South (>0.04) for early cohorts, followed by a sharp decline, a temporary recovery during the economic boom, and virtual disappearance after 1970. The consistently lower values of Q1Q4 compared to BHQ4 confirm that extreme upward mobility remains more challenging for children from the least educated families.

These mobility indicators fall into two categories: **inequality-invariant** indicators (1-COR, MU050, BHQ4, Q1Q4) measure only structural dependence between generations, while **inequality-sensitive** indicators (1-BETA, MIX) capture both intergenerational dependence and distributional changes across generations. This comprehensive approach reveals Italy's educational mobility exhibits significant regional variation with complex temporal dynamics.

Figure 1: Educational Mobility Indicators in Italy by Macro-Region and birth Cohorts



3.2 Corruption's Effect on Educational Mobility: Model

We investigate the relationship between corruption and intergenerational educational persistence through the following OLS specification:

$$\mathrm{E_i} = \beta_0 + \beta_1 \mathrm{E_i^P} + \beta_2 \mathrm{C_r} + \beta_3 \left(\mathrm{E}_i^{\mathrm{P}} \mathrm{x}\, \mathrm{C_r}\right) + \beta_4 X_i' + \alpha_{\mathrm{mp}} + \varepsilon_{\mathrm{i}} \qquad (1)$$

where $\mathrm{E_i}$ represents the child's years of education, $\mathrm{E_i^P}$ denotes the maximum years of education between parents, $\mathrm{C_r}$ is the standardized regional corruption index and X_i' is a vector of controls including regional socioeconomic indicators and their interactions with parental education, as well as quadratic terms for both parental education and corruption. Quadratic terms inclusion is justified by the possible correlation between these variables, which could lead to spurious interaction effects. If regional corruption and parental education are correlated (plausible if more educated parents tend to reside in less corrupt regions), the interaction term might inadvertently capture a quadratic effect of corruption rather than a true interaction effect without these controls.[1] Our specifications include fixed effects α_{mp}, for macro-area and period interactions to control for cohort-specific regional factors while avoiding multicollinearity with our regional corruption measure. These fixed effects capture the combined influence of geographical location (North, Center, South and Islands) and decennial periods where respondents were 16. The parameter of primary interest is $\boldsymbol{\beta_3}$: a positive value would indicate that corruption strengthens the intergenerational transmission of education, reducing educational mobility.

To explore potential non-linearities and threshold effects, we implement a quartile regression:

$$\mathrm{E_i} = \beta_0 + \textstyle\sum_{k=2}^{4} \beta_k\, \mathrm{Q_{ki}} + \beta_5\, \mathrm{E_i^P} + \beta_6 X_i' + \beta_7\, \mathrm{C_r} + \textstyle\sum_{k=2}^{4} \beta_{k+7}\, (\mathrm{Q_k}\ \mathrm{x}\ \mathrm{C})_{\mathrm{ir}} + \alpha_{\mathrm{mp}} + \varepsilon_{\mathrm{i}} \qquad (2)$$

where $\mathrm{Q_{ki}}$ represents parental education quartile indicators, allowing corruption's effect to vary distinctly across the educational distribution. The coefficient $\boldsymbol{\beta_7}$ captures corruption's baseline effect for children with parents in the lowest education quartile (reference category), while interaction coefficients $\boldsymbol{\beta_{k+7}}$ measure differential effects for higher quartiles.

[1] If regional corruption and parental education are correlated, we can express this as: $\mathrm{E_i^P} = \alpha \times \mathrm{C_r} + \mathrm{w_i}$. Thus, the interaction term decomposes as: $\alpha \times \mathrm{C_r}^2 + \mathrm{C_r} \times \mathrm{w_i}$

Our specification includes both quartile indicators and continuous parental education E_i^P, capturing threshold effects between major segments while accounting for within-quartile variation, recognizing that even within the same quartile, additional years of parental education may significantly impact children's educational attainment.

The model enables testing specific hypotheses about corruption's impact on education:

Uniform Disadvantage: If corruption depresses educational outcomes equally across all segments of family background, we would expect $\beta_7 < 0$ and β_{k+7} not statistically different from zero.

Social Calcification: If corruption reinforces educational inequalities by hindering children from less educated families, while offering protection for more privileged segments of society, we would expect $\beta_7 < 0$ while $\beta_{k+7} > 0$ and increasing in magnitude as k increases.

Protection Threshold: If even modest parental education provides substantial protection against corruption's negative effects, with limited additional benefits at higher levels, we would observe $\beta_7 < 0$, $\beta_9 > 0$, but β_{10} not statistically different from β_{11}. In this scenario, crossing a certain educational threshold provides insulation against corruption's adverse impacts, with diminishing returns beyond this point. If corruption disproportionately harms children from less educated families while having minimal or no effect on those from highly educated backgrounds, it effectively amplifies the intergenerational transmission of education.

To address potential endogeneity in our corruption measures, we implement a two-stage least squares (**2SLS**) approach throughout our analysis. Two sources of endogeneity may be identified: reverse causality, where entrenched low mobility may foster corruption-enabling institutions and omitted regional variables, such as social capital or familistic traditions simultaneously affecting both corruption and mobility patterns.

For continuous interaction model (model 1), we instrument the regional corruption variable, its interaction with parental education and the squared corruption term using standardized indices of cheating behavior and between-class variance in standardized testing, along with their interactions and squared values. Similarly, in our quartile-based specification (model 2), we instrument both the main effect of regional corruption and its quartile interactions using these base instruments and their corresponding quartile interactions. The propensity to cheat on standardized tests and to segregate students non-randomly are manifestations of the same underlying social norms that facilitate corruption, making them relevant instruments.

Some economic arguments may support the exclusion restriction for these instruments. First, our instruments are measured at the regional level, not at the individual or family level. While families' educational decisions are influenced by household income, parental education and local economic opportunities, it is implausible that the average rate of test cheating or between-class variance in a region would directly affect these family-level decisions except through their correlation with perceived corruption that, altering the value attributed to educational investments, shapes educational decision-making accordingly.

We acknowledge potential concerns regarding the exclusion restriction of our instruments given their educational origins. For example, in a region where students are systematically sorted into "good" and "less good" classes (high between-class variability), this practice might not only reflect corruption (which is what our instrument aims to capture) but, by reducing educational quality, it could also directly impact students' decisions about years of schooling. To address this potential violation of the exclusion restriction, we include educational quality as a control variable. It is worth highlighting that educational quality had already been included in our model, considering its documented correlation with educational mobility patterns (Van der Weide et al., 2024).

3.3 Corruption's Effect on Educational Mobility: Results

Table 1 presents our estimates from model 1, showing both OLS and IV specifications. The OLS results (column 1 and 2) reveal a significant negative effect of corruption on education, indicating that a one standard deviation increase in corruption is associated with approximately 1.5 fewer years of education. Parental education shows a robust positive effect. The interaction term between parental education and corruption is not statistically significant, suggesting that corruption's effect does not significantly vary by parental education level, or it does not mediate the educational persistence in this specification. The significant negative quadratic corruption effect in column 2 indicates that corruption's negative impact intensifies at higher corruption levels. Consistently with the literature on educational mobility, the interaction between parental education and regional controls is significantly negative, suggesting that better socioeconomic conditions (higher employment rates, higher income levels, lower income inequality, better educational quality, and higher separation rates) reduce educational persistence.

In the baseline IV model with no fixed effects (IV1), corruption shows a coefficient of -1.048 ($p < 0.01$), indicating that a one standard deviation increase in corruption reduces children's education by approximately 1.05 years. Crucially, the parental education - corruption interaction becomes statistically significant in the IV specifications, demonstrating that each additional year of parental education offsets corruption's negative impact by 0.053 years in regions with one standard deviation higher corruption. This finding confirms that corruption amplifies intergenerational persistence, reducing social mobility. Adding macro-area period fixed effects (IV2) yields similar results, with corruption's effect slightly reduced ($\beta = -1.022$, $p < 0.01$) and the interaction effect remaining significant ($\beta = 0.056$, $p < 0.1$). The inclusion of quadratic terms (IV3) shows a marginally smaller main effect ($\beta = -0.898$, $p < 0.05$) but a stronger interaction term ($\beta = 0.065$, $p < 0.05$). The quadratic corruption term is not significant ($\beta = -0.017$), suggesting the relationship is primarily linear rather than accelerating at higher corruption levels.All specifications confirm strong instrument relevance, with Kleibergen-Paap F statistics ranging from 97 to 248, well above conventional critical values. The Hansen J test p-values for the first two IV specifications (0.175 and 0.518) support valid overidentifying restrictions, while the lower p-value in the third specification (0.005) warrants caution when interpreting this more complex model.

All three IV specifications exhibit remarkable consistency, revealing two complementary interpreting ways: corruption strengthens the predictive power of parental education, or higher parental education provides increasing protection against corruption's negative effects. In either interpretation, the outcome remains the same - corruption calcifies educational hierarchies by disproportionately harming children from disadvantaged backgrounds while leaving those from educated families relatively insulated. The quantitative implications of these results are substantial. For instance, in the IV2 specification, a one standard deviation increase in corruption reduces educational attainment by approximately 1.02 years for children whose parents have no formal education. However, this negative effect diminishes by 0.056 years for each additional year of parental education. Consequently, **children whose parents completed primary education (5 years) experience a reduced negative impact of about 0.74 years** ($-1.022 + 5 \times 0.056$), while **those with university-educated parents (17 years) face a minimal reduction of only 0.07 years** ($-1.022 + 17 \times 0.056$). This pattern demonstrates that corruption not only lowers overall educational attainment but systematically magnifies existing inequalities, effectively converting temporary educational advantages into persistent intergenerational disparities.

Table 1: Effect of Corruption on Educational Persistence

Dep. Variable: Son's education (years)	**OLS1**	**OLS2**	**IV1**	**IV2**	**IV3**
Corruption	-1.543**	-1.492**	-1.048***	-1.022***	-0.898**
	(0.641)	(0.678)	(0.280)	(0.280)	(0.406)
Parental education	0.475***	0.500***	0.473***	0.475***	0.475***
	(0.008)	(0.035)	(0.008)	(0.008)	(0.008)
Parental education × Corruption	0.090	0.005	0.053*	0.056*	0.065**
	(0.063)	(0.010)	(0.029)	(0.029)	(0.029)
Regional controls (PC1)	0.295***	0.340***	0.328***	0.332***	0.335***
	(0.085)	(0.079)	(0.076)	(0.078)	(0.077)
Parental education × Regional controls	-0.011*	-0.013**	-0.011*	-0.012*	-0.014**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Corruption squared		-0.082**			-0.017
		(0.037)			(0.126)
Parental education squared		-0.001			-0.002
		(0.002)			(0.003)
Macro-area period fixed effects	Yes	Yes	No	Yes	Yes
Constant	8.306***	8.316***	8.502***	8.486***	8.458***
	(0.181)	(0.227)	(0.231)	(0.235)	(0.241)
Observations	6,539	6,539	6,539	6,539	6,539
R-squared	0.331	0.331	0.325	0.327	0.329
Kleibergen-Paap F statistic			97.136	98.718	248.269
Hansen J p-value			0.175	0.518	0.005

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2 presents our quartile regression results, revealing significant variation in corruption's impact across the parental education distribution. Focusing on the IV specification, the corruption coefficient (-0.787, $p < 0.01$) represents corruption's effect on children whose parents fall within the lowest education quartile. This negative effect progressively diminishes for higher quartiles, as shown by the positive and significant interaction terms: Q2 × Corruption (0.269, $p < 0.05$), Q3 × Corruption (0.494, $p < 0.01$), Q4 × Corruption (0.466, $p < 0.01$).
Net effects calculations demonstrate a clear gradient: a one SD increase in corruption reduces educational attainment by approximately 0.78 years for children with parents in the first quartile, 0.51 years for the second quartile, 0.29 years for the third quartile and 0.32 years for the fourth quartile. Equality tests between quartile interactions offer additional insights: the difference between second and third quartiles is marginally significant ($p = 0.095$), while differences between third and fourth quartiles are not significant ($p = 0.855$). This suggests moderate parental education (third quartile) provides substantial insulation against corruption's negative effects, with limited additional protection from higher education levels. These findings support our "**social calcification hypothesis**". The stark gradient across the parental education distribution shows corruption creates barriers disproportionately affecting children from less educated families, effectively hardening social hierarchies.

3.4 Underlying transmission mechanisms

Our empirical analysis demonstrates that corruption strengthens educational transmission. Two mechanisms may explain this pattern. First, **corruption appears to alter the valuation of education differently across social strata**. In corrupt environments, where advancement often depends on connections rather than merit, less-educated families may rationally reduce educational investments when observing that education yields limited returns locally. Highly educated families tend to provide role models that emphasize education's intrinsic value beyond immediate economic returns. These families maintain educational aspirations despite corruption, effectively neutralizing the devaluation effect that typically occurs in corrupt settings. Second, **corruption enhances the "inheritability" of status across generations**. When professional positions become effectively inherited through family networks, educational trajectories naturally align with anticipated occupational outcomes. Children pursue educational paths similar to their parents because these represent efficient routes to positions they're likely to inherit. To properly investigate these transmission mechanisms, understanding the relationship between corruption and occupational mobility becomes essential.

Table 2: Heterogeneity analysis results

Dep. Variable : Son's education (years)	OLS	IV
Corruption	-0.353***	0.787***
	(0.081)	(0.140)
Quartile 2 × Corruption	0.264***	0.269**
	(0.096)	(0.116)
Quartile 3 × Corruption	0.477***	0.494***
	(0.101)	(0.128)
Quartile 4 × Corruption	0.344***	0.466***
	(0.117)	(0.138)
Parental education: Quartile 2	0.635***	0.612***
	(0.176)	(0.177)
Parental education: Quartile 3	1.085***	1.029***
	(0.355)	(0.358)
Parental education: Quartile 4	1.062**	0.997*
	(0.514)	(0.518)
Parental education: years	0.383***	0.388***
	(0.039)	(0.040)
Regional controls (PC1)	0.197***	0.331***
	(0.073)	(0.077)
Macro-area period fixed effects	Yes	Yes
Observations	6,539	6,539
R-squared	0.334	0.330
Kleibergen-Paap F statistic		212.382
Hansen J p-value		0.109

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4. Occupational Mobility in Italy

4.1 Regional Mobility Patterns across Regions: Descriptive Analysis

Before analyzing corruption's relationship with occupational mobility, baseline mobility patterns across Italian macro-regions must first be established. This preliminary step is crucial: since corruption in Italy shows a clear north-south gradient (higher in the South), determining whether occupational mobility also varies significantly by region justifies the inclusion of macro-area fixed effects in subsequent corruption analysis.

Drawing on established methodology (Long and Ferrie, 2013), we construct contingency tables where rows represent offspring's occupational categories and columns represent parental occupational categories across eight occupational classifications: blue-collar workers, white-collar workers, teachers, middle managers, executives, professionals, entrepreneurs and self-employed/family business workers.

The **total mobility rate M** measures the percentage of individuals not following their parents' occupations (off the main diagonal of the mobility matrix). This intuitive measure can be misleading since differences may reflect varying occupational structures rather than actual mobility differences. To overcome this limitation, we build an **adjusted mobility measure M′** with Iterative Proportional Fitting (**IPF**), which standardizes distributions while preserving association patterns (for more details see section B1 in the appendix).

Since M and M′ don't allow for formal significance testing or identification of specific mobility barriers, we employ **odds ratio** analysis for more sophisticated assessment. In a 2×2 table, the odds ratio compares the relative chances of occupational inheritance versus mobility between different groups:

$$\theta = \frac{p_{11}p_{22}}{p_{12}p_{21}} \text{ or } \mathrm{OR} = \frac{\left(\frac{p_{11}}{p_{12}}\right)}{\left(\frac{p_{21}}{p_{22}}\right)}$$

It represents the ratio between the odds that a child of a parent in occupation 1 will have occupation 1 rather than occupation 2 and the odds that a child of a parent in occupation 2 will have occupation 1 rather than occupation 2. An odds ratio of 1 indicates independence, while values greater than 1 indicate that children tend to follow their parents' profession. For multi-category tables, we employ the odds ratio-based Altham statistic.

For two tables, P and Q, with r rows and s columns, the **Altham statistic $\mathbf{d(P, Q)}$** is defined as:

$$d(P, Q) = \left[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{l=1}^{r}\sum_{m=1}^{s}\left(\log \frac{p_{ij}p_{lm}}{p_{im}p_{lj}} - \log\frac{q_{ij}q_{lm}}{q_{im}q_{lj}}\right)^2\right]^{\frac{1}{2}}$$

It measures the distance between two mobility tables by summing the squared differences between the log odds ratios for all possible 2 x 2 subtables that can be extracted from the full contingency tables. In essence, it captures how differently the two tables depart from the condition of statistical independence. To assess mobility in a single table P, we compare it with table J representing perfect independence (maximum mobility). Larger values indicate stronger intergenerational association and thus lower mobility.[2]

The Altham statistic can be calculated more efficiently through analysis of variance (**ANOVA**) of the logarithms of cell frequencies p_{ij}, using the log-linear model:

$$\log(p_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

where μ is a general constant, α_i is the row effect (parental occupational category), β_j is the column effect (son occupational category), γ_{ij} is the interaction term capturing the specific association between parental category i and offspring category j. Under independence all γ_{ij} are zero. The Altham statistic turns out to be a function of the sum of the squared residuals (**RRS**) from the ANOVA[3] :

$$d(P, J) = \sqrt{\text{RSS x 4 x r x s}}$$

To test whether the association is statistically significant, we use a likelihood ratio test based on statistic $G^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{s} f_{ij}\log\left(\frac{f_{ij}}{e_{ij}}\right)$, where f_{ij} are the observed frequencies and e_{ij} are the expected frequencies under independence.[4]

---

[2] Since all elements in J equal 1, $d(P, J) = \left[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{l=1}^{r}\sum_{m=1}^{s}\left(\log \frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\right)^2\right]^{\frac{1}{2}}$

[3] For the detailed derivation of this relationship, refer to the Appendix, section B2.

[4] The G^2 statistic follows an asymptotic chi-square distribution with $r - 1$ and $s - 1$ degrees of freedom under the null hypothesis of independence.

The Altham statistic's **triangular property**, $d(P, Q) \leq d(P, J) + d(Q, J)$, allows us to construct a triangular ratio $\frac{d(P,Q)}{d(P,J) + d(Q,J)}$. When this ratio approaches 1, it indicates that societies exhibit qualitatively different mobility structures, meaning they have distinct patterns of which occupational transitions are fluid or rigid.

Table 3 presents the key mobility measures across Italy's three macro-areas.

Raw mobility rates M demonstrate a clear gradient, with the North exhibiting the highest mobility (61.12%), followed by the Center (58.64%) and the South/Islands (54.98%). However, when controlling for differences in occupational distributions through adjusted mobility rates M', these regional disparities largely disappear, with all regions converging around 58-59%. This suggests that much of the observed variation in raw mobility stems from different occupational structures rather than from inherent differences in mobility regimes.

Nevertheless, the Altham statistic reveals that the North has the lowest distance from independence ($d(P, J)$ = 73.38), indicating weaker intergenerational associations and fewer barriers to mobility, while the Center shows the highest value ($d(P, J)$ = 87.41).

Table 3: Mobility Measures by Geographical Macro-area

Macro-area	M	M'	d(P,J)	G^2
North	61.12%	58.06%	73.38	540.41***
Center	58.64%	58.64%	87.41	307.01***
South/Islands	54.98%	58.69%	83.42	561.98***

To assess whether the three macro-areas differ not only in mobility levels but also in mobility patterns, we computed distances between their mobility tables d(P,Q) and triangular ratios, as shown in Table 4. The remarkably high triangular ratios observed in our comparisons (North-Center: 0.61, North-South: 0.64, Center-South: 0.72) provide strong evidence that Italian regions differ not merely in mobility intensity but in their fundamental mobility structures. These elevated ratios, particularly between Center and South, indicate qualitatively distinct mobility regimes with different patterns of social barriers.

Table 4: Distances Between Regional Mobility Tables

Comparison	d(P,Q)	p-value	Triangular Ratio
North vs. Center	97.66	0.026	0.61
North vs. South/Islands	99.93	0.016	0.64
Center vs. South/Islands	123.47	0.368	0.72

Notes: The p-value assesses statistical significance of observed differences between mobility tables (d(P,Q)). The triangular ratio quantifies the qualitative nature of mobility differences. Thus, regions may exhibit qualitatively different mobility regimes (high triangular ratio) even when distances lack conventional statistical significance.

Decomposing the Altham statistic into its components is essential for identifying which occupational boundaries drive these regional patterns. Table 5 presents this decomposition, revealing the relevant transitional barriers that characterize each macro-area's mobility regime.

The North's predominant mobility barrier (component 1155, log odds ratio = 4.59) represents the contrast between blue-collar workers remaining in their class versus becoming executives, compared to executives' children following the same transitions. Quantitatively, this means that the odds of a blue-collar worker's child remaining in the blue-collar category rather than becoming an executive are 98.5 times higher than the odds of an executive's child becoming a blue-collar worker rather than remaining an executive. While this represents a substantial barrier, it is markedly lower than the corresponding barriers in the Center (log odds ratio = 5.79) and South (log odds ratio = 5.30 for similar transitions). The relative chance of upward mobility from blue-collar to executive positions in the North is more than three times greater than in the Center and twice as great as in the South. It suggests a more permeable class structure in northern Italy, potentially facilitated by stronger industrial base and more developed labor markets. The secondary barriers in the North are particularly revealing, as they uniquely involve teachers in contrast with entrepreneurial and self-employed categories. This pattern (components 3478 and 3378) suggests that in northern Italy, educational credentials may create distinct career paths that have limited crossover with entrepreneurial activities, revealing a specific form of occupational segmentation. Specifically, component 3478 indicates that the odds of a teacher's child becoming a middle manager rather than self-employed are 79.9 times higher than the corresponding odds for an entrepreneur's child. Similarly, component 3378 shows that the odds of a teacher's child remaining in teaching rather than becoming self-employed are 55.1 times higher than the analogous odds for an entrepreneur's child. These substantial barriers demonstrate pronounced occupational segregation between educational and entrepreneurial sectors in Northern Italy.

Central Italy exhibits not only the strongest barriers between blue-collar workers and executives (log odds ratio=5.79), but also a distinctive pattern involving executives, entrepreneurs and workers. The extremely high odds ratio for blue-collar to executive transitions indicates particularly rigid class boundaries in this region. Central Italy's unique feature is the pronounced separation between managerial careers in established organizations and entrepreneurial pathways, likely reflecting the region's dual economic structure: large bureaucratic organizations (especially in Rome with its extensive public administration) alongside traditional entrepreneurial sectors.

The South exhibits fundamentally different barriers centered on entrepreneurship. Blue-collar workers' children are 200.3 times more likely to remain blue-collar than become entrepreneurs compared to entrepreneurs' children making these transitions (component 1177, log odds ratio=5.30). This pattern, along with several other significant components involving entrepreneurial occupations (1175, 1176, each with log odds ratio = 4.63), confirms entrepreneurship as a closed category in southern Italy, where family businesses operate within tight networks restricting entry.

Italy presents **not a simple North-South mobility gradient** but a complex mosaic of **qualitatively different mobility regimes**, suggesting corruption may interact differently with these pre-existing regional patterns.

Table 5: Top Mobility Barrier Components by Macroarea

Macro-area	Component Code	Contribution to $d(P,J)^2$	% of Total	Log Odds Ratio	Odds Ratio
North	1155	21.07	1.57%	4.59	98.5
	3478	19.20	1.43%	4.38	79.9
	3378	16.06	1.19%	4.01	55.1
	3388	15.23	1.13%	3.90	49.4
	1156	14.39	1.07%	3.79	44.3
Center	1155	33.55	2.12%	5.79	327.2
	5185	28.05	1.77%	5.30	200.3
	5588	26.47	1.67%	5.14	170.6
	1558	19.88	1.26%	4.46	86.5
	2155	18.05	1.14%	4.25	70.1
South/Islands	1177	28.07	2.05%	5.30	200.3
	1175	21.47	1.57%	4.63	102.5
	1176	21.47	1.57%	4.63	102.5
	1135	21.47	1.57%	4.63	102.5
	1166	21.47	1.57%	4.63	102.5

Notes: Component codes indicate the specific occupational transitions being compared. The format XYZW represents a contrast between X→Y versus X→Z and W→Y versus W→Z transitions. The occupational categories are coded as: 1=Blue-collar worker, 2=White-collar worker, 3=Teacher, 4=Middle Manager, 5=Executive, 6=Professional, 7=Entrepreneur, 8=Self-employed/Family business. For example, component 1155 compares the transitions: Blue-collar→Blue-collar vs. Blue-collar→Executive with Executive→Blue-collar vs. Executive→Executive. The odds ratio of 98.5 indicates that the odds of a blue-collar worker's child remaining blue-collar versus becoming an executive are 98.5 times higher than the corresponding odds for an executive's child. Higher odds ratios indicate stronger barriers to mobility between the occupational categories involved in the contrast. The percentage contribution shows how much each component contributes to the overall Altham statistic for that area.

4.2 Corruption's Effect on Occupational Mobility and Education Devaluation: Models

This section develops an econometric framework to empirically test the hypothesized transmission mechanisms explaining corruption's effect on educational immobility. Our analysis examines two paths through which corruption may reinforce educational persistence: enhanced "inheritability" of occupational status and the devaluation of education as a mobility instrument. We first investigate whether corruption enhances occupational inheritance. Specifically, we test whether corruption creates systematic patterns where family background increasingly determines occupational outcomes. Subsequently, we assess if corruption reduces education's effectiveness as a mobility tool across different career pathways, thus evaluating whether rational agents might reduce educational investments in corrupt environments.

To examine the relationship between corruption and intergenerational occupational mobility, we employ a **multinomial logistic model**, given the categorical and non-ordered nature of our dependent variable, the occupational category of sons. The fundamental specification is:

$$\log\left(\frac{P(Y_i = j|\, X_i)}{P(Y_i = 1|\, X_i)}\right) = \alpha_j + \sum_{g=2}^{4} \beta_{jg}\, G_{ig} + \gamma_j C_r + \sum_{g=2}^{4} \delta_{jg}\left(G_{ig}\, x\, C_r\right) + \theta_j X_i' + \varphi_j W_r' \quad (3)$$

where Y_i represents the child's occupational category (with $Y_i = 1$ for manual occupations as reference category); G_{ig} are indicator variables for parental occupational categories (also with manual occupations as reference). We define parental occupation based on the parent with superior educational credentials, maintaining analytical continuity with our previous specifications examining educational persistence. Both parental and child occupations are classified into four aggregated categories: (1) blue-collar workers, (2) white-collar workers and teachers, (3) middle managers, executives and professionals and (4) entrepreneurs and self-employed/family business. C_r is the usual regional corruption index, X_i' contains individual years of education in linear and quadratic form and W_r' includes regional-level controls.

The crucial parameters are the interaction coefficients $\boldsymbol{\delta_{jg}}$, which capture how corruption modifies the effect of parental occupational background on children's occupational probabilities. When δ_{jg} is positive, corruption increases the relative advantage that children from parents in category g have in accessing occupation j compared to manual occupations.

This doesn't necessarily increase the absolute transition probability, as the effect also depends on how corruption influences other occupational alternatives. For meaningful interpretation, the following predicted probabilities of entering occupation j will be computed for each family background and occupational destination combination at different corruption levels:

$$P(Y_i = j|X_i) = \frac{\exp(Z_{ij})}{\sum_{k=1}^{4} \exp(Z_{ik})}, \text{with } \exp(Z_{i1}) = 1$$

$$Z_{ij} = \alpha_j + \sum_{g=2}^{4} \beta_{jg} G_{ig} + \gamma_j C_r + \sum_{g=2}^{4} \delta_{jg}(G_{ig} \text{ x } C_r) + X_i'\theta_j + W_r'\varphi_j$$

The marginal effect of corruption C_r on the probability that an individual i belongs to occupational category j is:

$$\frac{\partial P(Y_i = j|X_i)}{\partial C_r} = P(Y_i = j|X_i)\left[\left(\gamma_j + \sum_{g=2}^{4} \delta_{jg}G_{ig}\right) - \sum_{k=1}^{4} P(Y_i = k|X_i)\,(\gamma_k + \sum_{g=2}^{4} \delta_{kg}G_{ig})\right]$$

This expression can be decomposed into three components for more intuitive understanding:

Baseline Probability $P(Y_i = j|X_i)$: the odd that i belongs to occupational category j.

Direct Effect $\gamma_j + \sum_{g=2}^{4} \delta_{jg}G_{ig}$: includes γ_j, the baseline effect of corruption on category j, relative to the reference category and $\sum_{g=2}^{4} \delta_{jg}G_{ig}$, capturing how corruption changes the role of parental occupational background. For instance, if δ_{32} (the interaction between clerical parental background and managerial child outcomes) is positive, corruption particularly enhances the relative advantage that children of clerical workers have in accessing managerial positions compared to manual occupations.

Weighted Average Effect $\sum_{k=1}^{4} P(Y_i = k|X_i)\,(\gamma_k + \sum_{g=2}^{4} \delta_{kg}G_{ig})$: the weighted average of corruption effects across all occupational categories, where weights are the probabilities of belonging to each category. This term, subtracted from the previous component, explains why a positive coefficient does not necessarily imply a probability increase: if corruption more strongly enhances the relative advantage for other categories, the probability of j might decrease even with a positive coefficient.

We extend our analysis by incorporating fixed effects to control for unobserved factors that systematically vary across regions and time periods, which might simultaneously influence both corruption and mobility patterns. These fixed effects are macro-area-period interactions, with the period dummy based on when individuals made their occupational choices (assumed at age 25), allowing us to account precisely for labor market conditions specific to the time and place of career entry. The need for these fixed effects is demonstrated by the spatial heterogeneity in mobility patterns revealed in our previous descriptive analysis of occupational mobility.

Having established how corruption affects the inheritance of occupational status, we now turn to the second transmission mechanism: the potential devaluation of education as a mobility tool in corrupt environments. This requires extending our baseline model to explicitly capture how corruption moderates the relationship between educational attainment and occupational outcomes. We therefore augment model (3) by incorporating an interaction between individual years of education and regional corruption level:

$$\log\left(\frac{P(Y_i = j|X_i)}{P(Y_i = 1|X_i)}\right) = \alpha_j + \sum_{g=2}^{4} \beta_{jg} G_{ig} + \gamma_j C_r + \sum_{g=2}^{4} \delta_{jg}\left(G_{ig} \times C_r\right) + \eta_j E_i + \theta_j E_i^2 + \lambda_j (E_i \times C_r) + W_r' \phi_j \quad (4)$$

, where E_i represents years of education and $\boldsymbol{\lambda_j}$ captures how corruption modifies the marginal return to education for occupational outcome j relative to manual occupations (the reference category). For instance, a negative λ_j for entrepreneurial occupations would suggest that as corruption increases, the relative advantage that additional education provides for accessing entrepreneurial positions versus manual positions diminishes.
The predicted probability that an individual with education level E_i and living in a region with corruption level C_r enters occupation j is given by:

$$P(Y_i = j|X_i) = \frac{\exp(Z_{ij})}{\sum_{k=1}^{4} \exp(Z_{ik})}, \text{with } \exp(Z_{i1}) = 1$$

$$Z_{ij} = \alpha_j + \sum_{g=2}^{4} \beta_{jg} G_{ig} + \gamma_j C_r + \sum_{g=2}^{4} \delta_{jg}\left(G_{ig} \times C_r\right) + \eta_j E_i + \theta_j E_i^2 + \lambda_j (E_i \times C_r) + W_r' \phi_j$$

By calculating these probabilities at different years of education and corruption levels, we can visualize how the relationship between education and occupational attainment changes with corruption. The marginal effect of education on the probability of entering occupation j can be derived analytically as:

$$\frac{\partial P(Y_i = j|X_i)}{\partial E_i} = P(Y_i = j|X_i)\left[\eta_j + 2\theta_j E_i + \lambda_j C_r - \sum_{k=1}^{4} P(Y_i = k|X_i)(\eta_k + 2\theta_k E_i + \lambda_k C_r)\right]$$

It can be decomposed into three conceptual components:

Baseline Probability $P(Y_i = j|X_i)$: the odd that i belongs to occupational category j.

Direct Education Effect $\eta_j + 2\theta_j E_i + \lambda_j C_r$: includes the base effect of education (η_j), the derivative of the quadratic education term ($2\theta_j E_i$) allowing the marginal effect to vary with education level and, crucially, an interaction term ($\lambda_j C_r$) that modifies education's effect based on regional corruption levels.

Weighted Average Effect $\sum_{k=1}^{4} P(Y_i = k|X_i)(\eta_k + 2\theta_k E_i + \lambda_k C_r)$: represents the weighted average of education effects across all occupational categories, where weights are the probabilities of belonging to each category.

By simultaneously modeling the direct effect of parental occupation, its interaction with corruption and the education-corruption interaction, we can disentangle two mechanisms: (1) the "inheritance" effect through which corruption strengthens the parent-child occupational link regardless of education and (2) the education devaluation effect through which corruption alters the efficacy of education as a mobility tool differently across occupational pathways.

4.3 Corruption's Effect on Occupational Mobility and Education Devaluation: Results

Table 7 presents results from model 3. The dependent variable distinguishes four occupational categories, with manual occupations serving as the reference category. Each coefficient represents how the log-odds of entering a specific non-manual occupation relative to manual work changes with changes in the independent variables.

The upper panel reveals **strong occupational inheritance patterns**. All parental background coefficients are positive and significant, showing children from non-manual backgrounds have much higher odds of avoiding manual occupations. These effects follow clear patterns: children of managers/professionals show the strongest tendency to enter similar elite positions (coefficient 2.423, $p<0.01$), while entrepreneurial inheritance also demonstrates robust intergenerational transmission (coefficient 2.100, $p<0.01$).

Regional corruption alone shows no significant direct effects on occupational outcomes. However, the interaction terms reveal that corruption significantly amplifies advantages for sons from managerial/professional backgrounds, as evidenced by positive and significant coefficients (1.071, 1.307, and 1.229) across all non-manual categories. These values indicate that, ceteris paribus, in regions with one standard deviation higher corruption, the log-odds of children from elite backgrounds entering non-manual rather than manual occupations increase by 1.071 to 1.307 units, thus strengthening intergenerational persistence of occupational status.

Table 7 : The Effect of Corruption on Occupational Mobility

Dep. Variable: Son's Occupation	Clerical/ Educational	Managerial/ Professional	Entrepreneurial/ Self-employed
Parental occupation			
Clerical/Educational	1.333***	1.479***	0.630**
	(0.186)	(0.285)	(0.294)
Managerial/Professional	1.633***	2.423***	1.767***
	(0.406)	(0.464)	(0.466)
Entrepreneurial/Self-employed	0.951***	1.362***	2.100***
	(0.208)	(0.323)	(0.214)
Regional corruption	0.061	-0.207	0.208
	(0.139)	(0.259)	(0.165)
Interactions with corruption			
Clerical × Corruption	0.259	0.604*	0.154
	(0.220)	(0.311)	(0.319)
Managerial × Corruption	1.071*	1.307*	1.229*
	(0.639)	(0.693)	(0.679)
Entrepreneurial × Corruption	-0.108	-0.072	0.005
	(0.222)	(0.395)	(0.201)
Education (years)	0.711***	-0.068	-0.145
	(0.182)	(0.315)	(0.168)
Education^2	-0.012*	0.022*	0.010
	(0.007)	(0.012)	(0.007)
Observations	1,522		
Pseudo R^2	0.2204		
LR $chi^2(30)$	805.03***		

Notes: Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Reference category for the dependent variable: Manual occupations. Regional controls included.

Individuals from clerical backgrounds show selective benefits, with corruption enhancing their access to managerial positions (coefficient 0.604, $p<0.1$), increasing the log-odds by 0.604 units per standard deviation of corruption. Entrepreneurial parents' children show no significant interaction effects, suggesting their mobility remains unaffected by corruption. Education demonstrates expected patterns, with years of schooling strongly predicting entry into clerical/educational occupations rather than manual occupations (coefficient 0.711, $p<0.01$). The positive quadratic term for managerial outcomes (0.022, $p<0.1$) indicates that the effect of education on log-odds becomes stronger at higher educational levels, consistent with credentialism in elite occupations.

Table 8: Key Contrasts in Corruption Effects Across Occupational Transitions

Panel A: Predicted Probabilities for Key Occupational Transitions at different Corruption levels

Parental Occupation	**Son's Occupation**	**(-1 SD)**	**(+1 SD)**	**% Δ**
Managerial/ Professional	Manual	0.446	0.095	-79%
Managerial/ Professional	Clerical/Educational	0.325	0.447	+38%
Managerial/Professional	Managerial/Professional	0.147	0.178	+21%
Managerial/Professional	Entrepreneurial/Self-employed	0.083	0.281	+239%
Entrepreneurial/Self-employed	Entrepreneurial/Self-employed	0.251	0.358	+43%
Manual	Managerial/Professional	0.081	0.052	-36%

Panel B: Significant Contrasts in Marginal Effects of Corruption

Contrast	**Effect**	**p-value**
Managerial parents → Managerial vs. Manual parents → Manual	-0.166 (0.079)	0.037*
Entrepreneurial parents → Manual vs. Managerial parents → Manual	0.164 (0.082)	0.046*
Manual parents → Manual vs. Managerial parents → Manual	0.190 (0.084)	0.024*
Clerical parents → Manual vs. Managerial parents → Manual	0.207 (0.088)	0.019*
Clerical parents → Managerial vs. Managerial parents → Manual	0.191 (0.082)	0.020*
Managerial parents → Entrepreneurial vs. Managerial parents → Manual	0.271 (0.110)	0.014*
Entrepreneurial parents → Entrepreneurial vs. Managerial parents → Manual	0.231 (0.089)	0.010**

Notes: standard errors in parenthesis; ** $p < 0.01$, * $p < 0.05$. Effects represent differences in marginal effects of corruption between different occupational transitions.

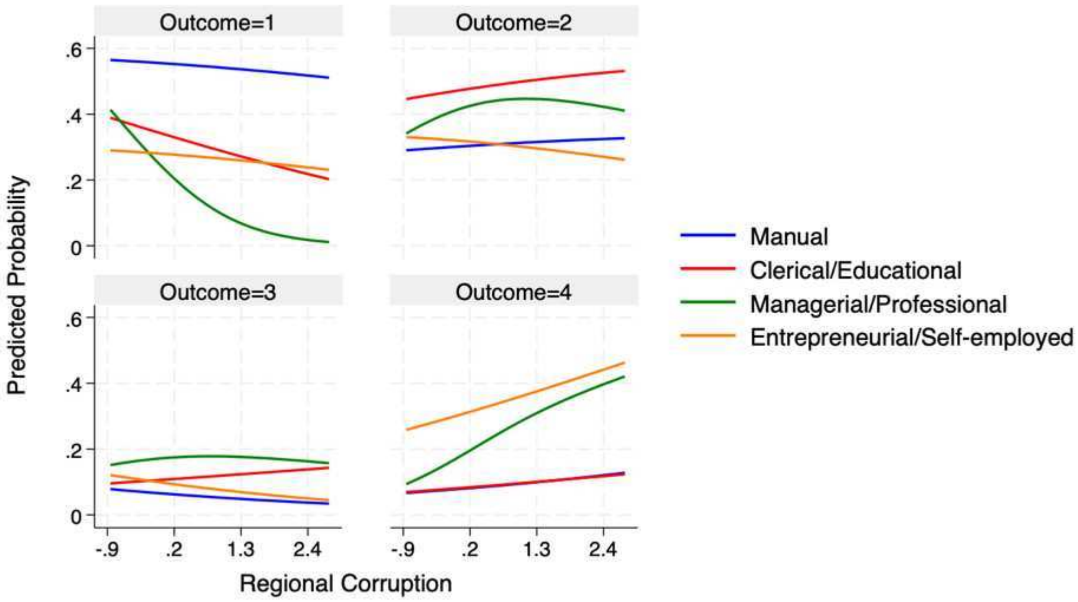
Table 8 quantifies how corruption reshapes occupational mobility pathways across generations. Panel A presents predicted probabilities for key occupational transitions at low versus high corruption levels (±1 standard deviation). Four distinct patterns emerge.

First, **children from managerial/professional backgrounds experience dramatic protection against downward mobility** in corrupt environments, with their probability of entering manual occupations plummeting from 44.6% to 9.5% (-79%). Second, rather than following identical career paths, elite children in corrupt contexts are redirected toward entrepreneurial/self-employed positions, with probabilities increasing from 8.3% to 28.1% (+239%). Third, children of entrepreneurial parents show enhanced occupational persistence when corruption is high, with inheritance probabilities rising from 25.1% to 35.8% (+43%). Fourth, **children from manual backgrounds face severely constrained upward mobility**, experiencing a 36% reduction in their probability of accessing managerial positions (from 8.1% to 5.2%).

Panel B provides formal statistical validation through contrasts in marginal effects, showing how corruption's impact differs across occupational transitions, controlling for education and regional characteristics. We calculated all 120 possible pairwise contrasts between the 16 parent-child occupational combinations, reporting only those reaching 5% statistical significance. Remarkably, only 7 contrasts achieved significance, with 6 of these sharing the same baseline: children of managers/professionals entering manual occupations. This pattern identifies elite downward mobility as the transition most sensitive to corruption. The contrast values quantify differential effects. For example, "Entrepreneurial parents → Entrepreneurial vs. Managerial parents → Manual" (0.231, $p<0.01$) indicates that when corruption increases by one standard deviation, the probability of entrepreneurial persistence increases by 23.1 percentage points more than the probability of elite downward mobility. Similarly, "Manual parents → Manual vs. Managerial parents → Manual" (0.190, $p<0.05$) shows that corruption reinforces manual occupational persistence relative to elite downward mobility. The single negative contrast (-0.166 for "Managerial parents → Managerial vs. Manual parents → Manual") indicates corruption's effect on manual persistence is relatively stronger than its effect on managerial persistence.

These results reveal corruption's primary function: creating a protective shield against elite downward mobility. While not necessarily promoting direct status reproduction, corruption most strongly prevents children from privileged backgrounds from experiencing downward mobility, thus calcifying social hierarchies. Figure 2 visualizes these findings, displaying predicted probabilities of entering different occupational categories based on parental occupation and regional corruption, while holding all control variables at their sample means.

Figure 2: Predicted Probabilities of Occupational Outcomes by Parental Background and Corruption Level



Notes: Four occupational outcomes: (1) Manual, (2) Clerical/Educational, (3) Managerial/Professional and (4) Entrepreneurial/Self-employed. Lines colour represents different parental backgrounds.

Notable patterns include: (i) The steep decline in the green line for Outcome 1 shows how corruption virtually eliminates the risk of downward mobility into manual work for children of managers/professionals, dropping from about 40% at -1 SD to near 0% at +2.5 SD. (ii) The sharp rise in the green line for Outcome 4 demonstrates how these same elite children are redirected toward entrepreneurial careers in corrupt environments, increasing from about 10% to over 40%. (iii) The downward slope of the blue line for Outcome 3 illustrates how corruption reduces upward mobility opportunities for children from manual backgrounds into managerial positions. (iv) The increasing orange line in Outcome 4 shows how children of entrepreneurs become increasingly likely to maintain entrepreneurial status as corruption rises.

We addressed unobserved regional and temporal heterogeneity by including macro-area and period fixed effects (Table A2, Appendix). Our results remain robust: the interaction between managerial background and corruption stays positive and marginally significant for both managerial (1.174, $p<0.1$) and entrepreneurial outcomes (1.142, $p<0.1$), confirming elite protection against downward mobility. The clerical background-corruption interaction for managerial outcomes (0.608, $p<0.1$) also persists, preserving the selective upward mobility pattern. Notably, corruption's direct effect on entrepreneurial occupations becomes significantly positive (0.303, $p<0.1$), suggesting its channeling effect toward entrepreneurship strengthens when controlling for regional heterogeneity. The adoption of macro-area fixed effects might not fully capture persistent regional patterns correlated with corruption, potentially leaving some regional heterogeneity unaccounted for. However, when we implement the exact same model specification on a more extensive sample including older cohorts, whose occupational decisions predate our corruption measures, all corruption coefficients become statistically insignificant. If our results were merely capturing unobserved persistent regional characteristics, we would expect to find similar patterns across all cohorts.

Results from model (4) are presented in Table 10. For managerial/professional outcomes, the education-corruption interaction coefficient is positive and marginally significant (0.1, $p<0.1$), indicating that in regions with one standard deviation higher corruption, the log-odds of entering managerial rather than manual occupations increase by an additional 0.1 units for each year of education. This may suggest formal credentials gain importance as competence signals when merit-based selection is compromised. For entrepreneurial outcomes, the interaction is negative and significant (-0.062, $p<0.05$), meaning that in regions with one standard deviation higher corruption, the log-odds of entering entrepreneurial versus manual occupations decrease by 0.062 units per year of education. This suggests **entrepreneurial success in corrupt environments depends more on informal networks than formal qualifications**.

The interaction for clerical/educational outcomes is not statistically significant (-0.018). This may reflect the standardized hiring requirements for educational positions, where corruption may influence who ultimately obtains positions but does not alter the formal educational qualifications for eligibility. A joint test of all education-corruption interactions ($\chi^2(3)=8.90$, $p=0.0307$) confirms corruption systematically alters educational returns across occupations. Thus, corruption reshapes education's value as a mobility instrument differently across occupational pathways, beyond its effects through family background channels.

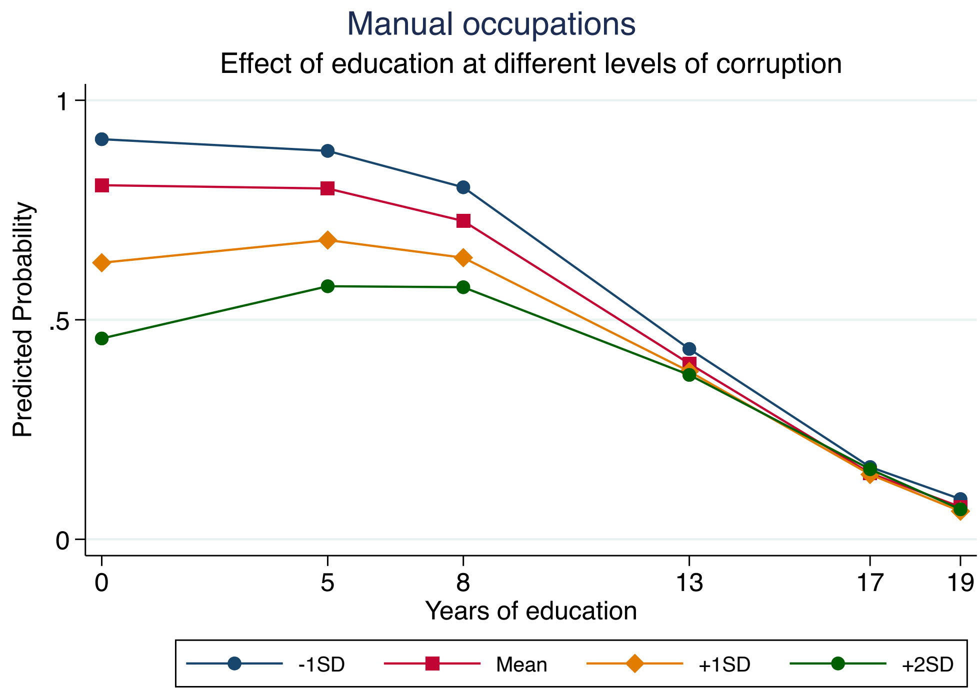
Table 10: Testing the Educational Devaluation Channel

Variable	Clerical/ Educational	Managerial/ Professional	Entrepreneurial/ Self-employed
Parental Occupation			
Clerical/Educational	1.338***	1.485***	0.630**
	(0.187)	(0.287)	(0.295)
Managerial/Professional	1.615***	2.434***	1.737***
	(0.410)	(0.470)	(0.470)
Entrepreneurial/Self-employed	0.954***	1.355***	2.101***
	(0.208)	(0.332)	(0.216)
Regional Corruption	0.198	-1.333	0.950**
	(0.425)	(0.936)	(0.451)
Interactions with Corruption			
Clerical × Corruption	0.288	0.606*	0.312
	(0.217)	(0.338)	(0.334)
Managerial × Corruption	1.078*	1.185	1.356*
	(0.647)	(0.722)	(0.699)
Entrepreneurial × Corruption	-0.154	-0.216	0.033
	(0.218)	(0.461)	(0.204)
Education (years)	0.729***	-0.019	-0.046
	(0.186)	(0.320)	(0.187)
Education × Corruption	-0.018	0.100*	-0.062**
	(0.027)	(0.058)	(0.030)
Observations	1,522		
Pseudo R^2	0.2243		
LR $chi^2(36)$	819.27***		

Notes: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Reference category for the dependent variable: Manual occupations. Joint test of education-corruption interaction terms: $\chi^2(3) = 8.90$, $p = 0.0307$. Regional controls and squared corruption and education terms included but not reported.

Figure 3 visualizes how the relationship between education and manual occupation probability varies across corruption levels. The predicted probabilities are calculated at specific education levels (0, 5, 8, 13, 17, 19 years) and corruption intensities (-1 SD, mean, +1 SD, +2 SD), while maintaining all other variables at their observed values. This average marginal effects approach averages the predicted probabilities across the actual distribution of parental occupations and other characteristics in the sample, providing a population-averaged view of how corruption moderates education's effect on occupational outcomes.

Figure 3: The Education effect on the Probability of entring Manual Occupation at Different Corruption Levels



In low-corruption environments (-1 SD), education produces a steady decline in manual occupation probability. In high-corruption contexts (+2 SD), the relationship becomes non-monotonic: manual occupation probability initially increases with early education, then decreases at a much slower rate compared to low-corruption regions. This distinct pattern between corruption levels is most pronounced at intermediate education levels (5-13 years), while curves eventually converge at university education (17-19 years), with probabilities between 0.05-0.15. The convergence suggests that advanced studies retain their value even in corrupt environments, neutralizing corruption's detrimental effect on educational returns. This visualization confirms our regression findings that **corruption alters education's effectiveness as a mobility tool, with particularly detrimental effects for those with primary and secondary education, while university education maintains its protective power.**

5. Intergenerational Income Mobility

Having quantified both educational and occupational mobility, our analysis now extends to intergenerational income mobility to provide a comprehensive assessment of economic status transmission. Income mobility represents the most direct measure of economic opportunity across generations and serves as the primary metric through which the cumulative effects of educational and occupational inheritance manifest.

5.1 TSTSLS Methodology and Bias Correction

The fundamental challenge in estimating intergenerational income mobility stems from the absence of directly linked parent-child income data in most survey datasets. The **TSTSLS** methodology provides an innovative solution by leveraging two independent samples with complementary information. The methodology begins with the estimation of a **Mincer-type earnings equation** using an auxiliary sample of potential fathers ("**pseudo-fathers**") from the earliest available SHIW waves:

$$\log\left(y^{pf}\right) = \omega' z^{pf} + v$$

where y^{pf} denotes the income of "pseudo-fathers" in the auxiliary sample, z^{pf}represents a vector of their socio-demographic characteristics including education, occupation, economic sector and geographic location (specifically, a southern Italy indicator) and v captures unobserved determinants of income.

The second stage applies the estimated coefficients $\widehat{\omega}'$ to the characteristics that sons report about their fathers in the most recent waves, generating predicted log paternal incomes:

$$\widehat{\log\left(y^{f}\right)} = \widehat{\omega'}\, z^{f}$$

The intergenerational income elasticity (**IGE**) is then estimated through:

$$\log\left(y^{s}\right) = \alpha + \beta\, \widehat{\log\left(y^{f}\right)} + \varepsilon$$

where y^{s} represents the son's income and $\boldsymbol{\beta}$ is the **intergenerational transmission parameter**.

The TSTSLS estimator suffers from two systematic biases that require correction and can be understood through the following decomposition (we omit log notation for clarity):

$$\text{plim } \hat{\beta}^{\text{TSTSLS}} = \frac{\text{cov}(y^s, \hat{y}^f)}{\text{var}(\widehat{y^f})} = \frac{\text{cov}(y^s,\ y^f + \hat{y}^f - y^f)}{\text{var }(\widehat{y^f})}$$
$$= \frac{\text{cov}(y^s,\ y^f)}{\text{var}(y^f)} \frac{\text{var}(y^f)}{\text{var}(\hat{y}^f)} - \frac{\text{cov}(y^s,\ y^f - \hat{y}^f)}{\text{var}(y^f)} \frac{\text{var}(y^f)}{\text{var}(\hat{y}^f)}$$

It follows that:

$$\text{plim } \hat{\beta}^{\text{TSTSLS}} = \frac{(\text{plim}\hat{\beta} - \eta)}{x}, \text{ with } \eta \equiv \frac{\text{cov}(y^s, y^f - \hat{y}^f)}{\text{var}(y^f)} \text{ and } x \equiv \frac{\text{var}(\hat{y}^f)}{\text{var}(y^f)}$$

where $\hat{\beta}$ represents the true IGE that would be obtained with directly linked data, $\boldsymbol{\eta}$ denotes the projection bias and **x** the variance bias.

The **variance bias** arises because predicted incomes exhibit less dispersion than actual incomes. Since OLS predictions minimize squared errors, we have $\text{var }(\hat{y}^f) \leq \text{var}(\ y^f)$, implying $x\ \leq 1$. This compression inflates the estimated intergenerational elasticity. Following Orihuela et al. (2023), we implement a variance correction by augmenting predicted incomes with a stochastic component. The variance corrected prediction is $\hat{y}^{fv} = \widehat{\omega'}\ z^f + \tilde{e}$, where $\tilde{e} \sim N(0, \sigma_v^2)$ with σ_v^2 being the estimated variance of residuals from the first stage regression. This correction ensures that $\text{var}(\hat{y}^{fv}) = \text{var}(y^f)$, effectively eliminating the variance bias.[5]

The **projection bias** arises from a fundamental limitation of the TSTSLS approach: observable characteristics used in the first-stage regression (education, occupation, sector, location) capture only part of what determines income. Individual ability, social capital, family networks and motivation also influence earnings but remain unmeasured in our Mincer equation. This creates a systematic bias when these unobserved factors are intergenerationally transmitted. For

[5] This equality holds under the assumption that the variance of unobservable actual fathers' incomes equals the variance of incomes in the pseudo-fathers sample, i.e., $\text{var}(y^f) = \text{var}(y^{pf})$. This assumption is reasonable when both samples are drawn from the same population. To ensure this, the true fathers' birth years are used to match individuals in the younger sample with the appropriate potential-father cohort in the older sample. The variance correction works because: (i) the predicted income $\widehat{\omega'}\ z^f$ has variance less than the true income due to OLS properties; (ii) adding an error term with variance σ_v^2 (the residual variance from the first-stage regression) to the prediction (iii) results in $\text{var}(\hat{y}^{fv}) = \text{var}(\widehat{\omega'}\ z^f) + \sigma_v^2 = \text{var}(y^{pf}) = \text{var}(y^f)$.

instance, if high-ability fathers tend to have high-ability children and ability affects income beyond what education and occupation can explain, then the prediction error from our first-stage regression will correlate with children's incomes leading to $\mathrm{cov}\left(y^s,\ y^f - \hat{y}^f\right) \neq 0$. To address this bias, we follow Orihuela et al. (2023) assuming that the correlation between prediction errors and children's incomes is proportional to the correlation between predicted incomes and son's incomes: $\mathrm{cov}\left(y^s, y^f - \hat{y}^f\right) = \lambda\, \mathrm{cov}\,(y^s, \hat{y}^f)$. This assumption reflects the intuition that the stronger the relationship between predicted fathers' incomes and son's incomes, the more likely it is that omitted variables driving the prediction errors also affect children's outcomes. Under this assumption, the double-corrected estimator is (see the appendix for complete derivation):

$$\hat{\beta}^{DC-TSTSLS} = \hat{\beta}^{VC-TSTSTLS}(1+\lambda)$$

Since λ cannot be directly estimated, we employ a theory-driven lower bound $\lambda_L = 2\,(1-R^2)^3$, where R^2 is the coefficient of determination from the first-stage regression. This functional form has several appealing properties: (i) when $R^2 = 1$, $\lambda_L = 0$, indicating no projection bias. (ii) As R^2decreases, λ_L increases at an accelerating rate, reflecting greater potential for unobserved factors to create bias. (iii) The cubic term ensures conservative correction for moderate R^2 values, while providing substantial adjustment when much variance remains unexplained (see the appendix for more details). The final double-corrected estimator is $\hat{\beta}^{DC-TSTSLS} = \hat{\beta}^{VC-TSTSTLS}(1+\lambda_l)$. It provides an accurate estimate of the true intergenerational elasticity, accounting for the systematic underestimation caused by unobserved intergenerationally transmitted factors.

As a complementary approach to the traditional IGE, we estimate intergenerational mobility using the **rank-rank slope** methodology following recent developments in the literature (Chetty et al., 2014; Corak, 2017; Cortes Orihuela et al., 2023). The rank-rank specification estimates:

$$r^s = \alpha_{rr} + \beta_{rr}\hat{r}^f + \varepsilon$$

where r^s represents the percentile rank of children in their income distribution and $\hat{r}^f$ represents the percentile rank of fathers based on their TSTSLS-imputed incomes. The rank-rank approach offers an important advantage over the log-linear IGE specification, particularly in the context of TSTSLS estimation: it does not suffer from the variance bias that plagues the traditional IGE. This property arises because the rank transformation preserves relative positions while fixing the variance of both actual and imputed rankings at $\frac{1}{12}$ (for uniformly distributed ranks between 0 and 1). The estimated coefficient is:

$$\widehat{\beta_{rr}}^{TSTSLS} = \frac{\mathrm{cov}\left(r^s, \widehat{r^f}\right)}{\mathrm{var}\left(\widehat{r^f}\right)}$$

To analyze the bias, we note that the true rank of fathers can be written as: $r^f = \widehat{r^f} + u$, where u is the rank prediction error. Therefore, exploiting that both true and imputed ranks are uniformly distributed:

$$\mathrm{plim}\ \widehat{\beta_{rr}}^{TSTSLS} = 12\ \mathrm{cov}\left(r^s, r^f\right) - 12\ \mathrm{cov}\ (r^s, u)$$

Since the true rank-rank coefficient is: $\widehat{\beta_{rr}} = 12\ \mathrm{cov}\left(r^s, r^f\right)$, we can therefore write:

$$\mathrm{plim}\ \widehat{\beta_{rr}}^{TSTSLS} = \widehat{\beta_{rr}} - 12\ \mathrm{cov}\ (r^s, u)$$

Fathers who have higher actual ranks than predicted (u>0) typically possess advantageous unobserved characteristics such as innate ability, social networks or family wealth. Since these characteristics are often partially transmitted to children, we expect children of "underpredicted" fathers to also rank higher in their generation. This creates a positive covariance between children's ranks and fathers' prediction errors: $\mathrm{cov}\ (r^s, u) > 0$. Because the Rank-Rank TSTSLS estimator subtracts this positive covariance from the true rank-rank correlation, it produces a conservative estimate. Empirically, this downward bias is relatively modest, making the rank-rank approach more reliable for comparative mobility studies.

5.2 Intergenerational Income Persistence: Empirical Estimates for Italy

Before presenting our TSTSLS estimates, we examine the characteristics of our samples to understand both intergenerational changes in Italy and potential challenges in the TSTSLS methodology (see Table A3 and Table A4 in the appendix). Table A3 compares sons and their actual fathers at the same life-cycle stage [6], revealing substantial structural changes between generations. Educational attainment has improved dramatically: the share with upper secondary education increased from 11.56% among fathers to 40.59% among sons, while university degree holders rose from 3.59% to 18.45% (including post-graduate degrees). The occupational structure has also transformed, with manual workers declining by 20.24 percentage points and managerial and professional positions collectively increasing by 21.36 percentage points.

Table A4 highlights **systematic differences between actual fathers** (reported by sons) **and pseudo-fathers** (directly observed in earlier SHIW waves). Consistent with findings by both Mocetti (2007) and Piraino (2007), fathers described by sons appear systematically less educated, more concentrated in manual occupations and more represented in agriculture compared to directly observed pseudo-fathers. Several factors explain this pattern. First, despite restricting our sample to sons whose fathers were born after 1939 to ensure overlap with the pseudo-father cohort (1939 -1958), a five-year mean age gap persists (1942.7 vs. 1947.9), placing actual fathers in a historical context with more limited educational and occupational opportunities. Further restriction to eliminate this gap would excessively reduce our sample size. Second, following Piraino's (2007) explanation of "differential fertility", less educated parents with manual occupations tend to have more children, resulting in their overrepresentation in the actual father sample. Third, recall bias in retrospective data, well-documented in the intergenerational literature, may lead sons to systematically underestimate their fathers' educational and occupational achievements.

These systematic differences do not compromise TSTSLS validity. As Solon and Inoue (2005) emphasize, the procedure "**implicitly corrects for differences in the distribution of variables between the two samples**", since the first stage estimates returns to characteristics in the fathers' period, which are then applied to the actual characteristics reported for fathers, regardless of their distribution. The persistence of this across different SHIW waves, as

[6] Respondents were asked to provide information about their fathers when they were the same age as the respondents are currently.

documented in both Mocetti (2007) and Piraino (2007), confirms it represents an empirical regularity in Italian retrospective data rather than a limitation specific to our particular survey years. Results from the first stage regression, for each income specification are reported in the appendix (Table A5).

Our TSTSLS estimates of intergenerational income mobility in Italy (Table 11) demonstrate the critical importance of proper bias correction in TSTSLS estimation. Previous uncorrected TSTSLS estimates from studies using the same SHIW data placed Italy's IGE between 0.43-0.65 (Piraino, 2007; Mocetti, 2007), suggesting lower mobility than the United States (IGE: 0.34-0.53, Solon, 1992; Bjorklund and Jantti, 1997) or United Kingdom (0.58, Dearden et al., 1997). The uncorrected TSTSLS estimate in our analysis yields an implausibly high IGE of 0.650, consistent with these previous Italian studies. After applying the variance correction methodology, our VC-TSTSLS estimate drops dramatically to 0.222 (SE = 0.046), with a further increase to **0.376** after correcting for projection bias using $\lambda = 2(1-R^2)^3$, where $R^2 = 0.2978$ from our first-stage regression. These double-corrected estimates align more reasonably with Acciari et al.'s (2022) **gold standard of 0.25** derived from administrative tax data. **About one quarter of income advantages are transmitted across generations according to their administrative data, while our estimate indicates closer to three-eighths (37.6%) transmission**.

Table 11: TSTSLS Estimates of Intergenerational Income Mobility in Italy

	TSTSLS	Bootstrap SE (TSTSLS)	VC-TSTSLS	Bootstrap SE (VC-TSTSLS)	DC-TSTSLS
IGE					
Net Total Income	0.650***	0.088	0.222***	0.046	0.375
Net Income excluding Financial Capital	0.655***	0.102	0.215***	0.043	0.367
Labor Income	0.644***	0.087	0.190***	0.050	0.331
RRS					
Net Total Income	0.362***	0.033			

Notes: Bootstrap standard errors based on 1,000 replications. The DC-TSTSLS estimates are derived as transformations of VC-TSTSLS, thus do not have their own standard errors from regression output.
*** $p < 0.001$

The persistence of economic status varies significantly across income specifications, with a consistent pattern showing decreasing persistence as we move from broader to narrower income definitions. The double-corrected IGE for net income excluding financial capital (0.367) is lower than that for total net income (0.376), while labor income alone yields the lowest DC-TSTSLS estimate (0.331). This pattern aligns with theoretical expectations that labor income should be less persistent across generations than broader income measures that include capital income, further validating our methodological approach.

Our rank-rank slope analysis provides additional insights while avoiding the variance bias issues inherent in traditional IGE estimation. The national **RRS of 0.362** (SE = 0.033) exceeds Acciari et al.'s baseline estimate of **0.246**, but importantly, when they adjust for tax evasion—a crucial consideration given documented rates of 40% for self-employment income and 80% for rental income in Italy—their RRS increases to **0.286**. Under perfect mobility, the RRS would equal 0, indicating no relationship between parent and child ranks. The quantitative implications of our RRS estimate are substantial: it implies that **children from families at the 90th percentile of the income distribution rank approximately 29 percentiles higher than those from 10th percentile families** (0.362 × 80 = 28.96).

Our corrected IGE of 0.376 and RRS of 0.362 position Italy closer to France (IGE 0.41, Lefranc and Trannoy, 2005) and within the range of other developed economies. While direct comparisons are limited by methodological differences, our results align with the expectation that Southern European countries would exhibit mobility patterns between the high-mobility Nordic nations (Sweden IGE 0.13-0.28, Osterberg, 2000; Bjorklund and Jantti, 1997) and lower-mobility Anglo-Saxon countries.Although differences persist between our estimates and the Acciari et al. benchmark, these can be attributed to different datasets and historical periods covered. Nevertheless, it is important to acknowledge that no estimation strategy using imputed incomes can ever be as accurate as analyses based on directly linked parent-child income data, which remain the gold standard for income mobility research.

6. Conclusions

This thesis demonstrates that Italy exhibits low intergenerational mobility across all dimensions. Educational mobility shows strong regional persistence, with Southern regions consistently exhibiting lower mobility than Northern areas; occupational mobility displays qualitatively different regimes across macro-regions, from educational-entrepreneurial segregation in the North to closed entrepreneurial networks in the South; income mobility estimates indicate 38% of income advantages transmit across generations, with children from top-decile families ranking 29 percentiles higher than those from bottom-decile families.

Corruption emerges as a powerful amplifier of these mobility constraints. We examine corruption's effects using behavioral instruments that Italy's National Anti-Corruption Authority (ANAC), recognizes as valuable for quantifying corruption risk. A one standard deviation increase in corruption reduces educational attainment by 0.79 years for children from the lowest education quartile, while providing substantial protection for elite offspring. In occupational outcomes, corruption creates a dramatic "elite protection effect" - reducing downward mobility risk by 79% for children from managerial backgrounds while cutting manual workers' children's access to elite positions by 36%.

Our results carry significant policy implications: anti-corruption initiatives should be conceptualized as first-order mobility policies; educational interventions should specifically target vulnerable groups; hiring processes should emphasize transparency to weaken family background's influence. Several promising directions for future research emerge: examining corruption's differential effects across public and private sectors and conducting longitudinal studies to track how specific anti-corruption reforms affect mobility patterns over time, revealing which institutional improvements most effectively restore intergenerational mobility.

By bridging corruption studies and intergenerational mobility research, we demonstrate that corruption's consequences extend beyond economic efficiency to what ANAC identifies as "the cohesion of the community, trust, loyalty and reciprocal trust among social actors," which forms the basis for effective economic transactions and citizen-institution relationships. Our findings confirm ANAC's insight that corruption fundamentally alters the social fabric, transforming temporary advantages into permanent privileges that systematically undermine meritocratic principles of social advancement.

References

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante. 2022. "And Yet It Moves: Intergenerational Mobility in Italy." American Economic Journal: Applied Economics 14, no. 3: 118-163.

Björklund, Anders, and Markus Jäntti. 1997. "Intergenerational Income Mobility in Sweden Compared to the United States." American Economic Review 87, no. 5: 1009-1018.

Borcan, Oana, Mikael Lindahl, and Andreea Mitrut. 2017. "Fighting Corruption in Education: Evidence from a Field Experiment in Romania." Journal of Public Economics 150: 25-38.

Checchi, Daniele, Andrea Ichino, and Aldo Rustichini. 1999. "More Equal but Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the US." Journal of Public Economics 74, no. 3: 351-393.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics 129, no. 4: 1553-1623.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." Journal of Economic Perspectives 27, no. 3: 79-102.

Cortes Orihuela, Joaquín, Pablo Pérez Akaki, and Roy van der Weide. 2023. "Everything's Not Lost: Revisiting TSTSLS Estimates of Intergenerational Mobility in Developing Countries." International Tax and Public Finance 30, no. 4: 892-920.

Cortes Orihuela, Joaquín, Pablo Pérez Akaki, and Roy van der Weide. 2024. "The Two-Sample Two-Stage Least Squares Method to Estimate the Intergenerational Earnings Elasticity." Journal of Economic Inequality 22, no. 1: 45-68.

Erikson, Robert, and John H. Goldthorpe. 1992. The Constant Flux: A Study of Class Mobility in Industrial Societies. Oxford: Oxford University Press.

Golden, Miriam, and Lucio Picci. 2005. "Proposal for a New Measure of Corruption, Illustrated with Italian Data." Economics & Politics 17, no. 1: 37-75.

Grawe, Nathan D. 2006. "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence." Labour Economics 13, no. 5: 551-570.

Lefranc, Arnaud, and Alain Trannoy. 2005. "Intergenerational Earnings Mobility in France: Is France More Mobile than the US?" Annales d'Économie et de Statistique 78: 57-77.

Lindner, Attila. 2014. "Does Corruption Reduce Returns to Education? Evidence from China." Economics of Education Review 41: 84-96.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." American Economic Review 103, no. 4: 1109-1137.

Mocetti, Salvatore. 2007. "Intergenerational Earnings Mobility in Italy." B.E. Journal of Economic Analysis & Policy 7, no. 2: 1-30.

Mocetti, Salvatore, and Tommaso Orlando. 2017. "Corruption and Personnel Selection in the Public Sector." Bank of Italy Occasional Paper No. 402.

Nifo, Annamaria, and Gaetano Vecchione. 2014. "Do Institutions Play a Role in Skilled Migration? The Case of Italy." Regional Studies 48, no. 10: 1628-1649.

Piraino, Patrizio. 2007. "Comparable Estimates of Intergenerational Income Mobility in Italy." B.E. Journal of Economic Analysis & Policy 7, no. 2: 1-27.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." American Economic Review 82, no. 3: 393-408.

Tabellini, Guido. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." Journal of the European Economic Association 8, no. 4: 677-716.

van der Weide, Roy, Christoph Lakner, Daniel Gerszon Mahler, Ambar Narayan, and Rakesh Ramasubbaiah. 2024. "Intergenerational Mobility around the World: A New Database." Journal of Development Economics 166: 103849.

Appendices

Appendix A: Supplementary Tables

Table A1: Descriptive Statistics

Panel A: Summary Statistics for Mobility Analysis Samples

	Educational Mobility Sample	Occupational Mobility Sample
	Mean (SD) or %	Mean (SD) or %
Sample size	36,883	6,506
Age	57.88 (15.64)	43.83 (6.82)
Birth year	1954.04 (15.98)	1963.85 (7.24)
Age range	21-104	30-55
Male	58.7%	65.6%
Educational attainment		
Years of education - sons	10.33 (4.67)	11.70 (3.77)
Years of education - parents	5.99 (4.48)	6.84 (4.09)
Occupational distribution – sons		
Manual workers	42.1%	39.4%
Clerical/Educational	30.9%	34.4%
Managerial/Professional	11.5%	11.8%
Entrepreneurial/Self-employed	15.5%	14.4%
Occupational distribution - parents		
Manual workers	57.6%	54.4%
Clerical/Educational	14.6%	18.7%
Managerial/Professional	6.0%	9.5%
Entrepreneurial/Self-employed	21.9%	17.4%
Geographic distribution		
North	38.1%	43.6%
Center	19.0%	20.2%
South/Islands	42.9%	36.3%

Panel B : Descriptive Statistics of Indicators by Macro-area

	North		Center		South		Italy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Corruption	0.113	0.059	0.087	0.053	0.374	0.170	0.212	0.176
Cheating	0.016	0.008	0.032	0.013	0.059	0.022	0.037	0.025
Class Variability	45.06	7.072	51.366	5.983	62.463	8.823	53.281	10.840
Employment	0.673	0.032	0.619	0.046	0.504	0.043	0.595	0.087
Income	16496.7	676.6	14141.7	953.7	10698	1111.5	13706	2809.9
Income Inequality	0.405	0.015	0.409	0.012	0.434	0.015	0.417	0.019
Educational quality	0.735	0.090	0.645	0.029	0.391	0.173	0.579	0.201
Separation Rate	1.501	0.294	1.243	0.149	0.704	0.163	1.131	0.427

Table A2: The Effect of Corruption on Occupational Mobility with Fixed Effects

Dep. Variable: Son's Occupatioal Category	**Clerical/ Educational**	**Managerial/ Professional**	**Entrepreneurial/ Self-employed**
Parental Occupational Category			
Clerical/Educational	1.338***	1.485***	0.588**
	(0.187)	(0.287)	(0.296)
Managerial/Professional	1.614***	2.455***	1.717***
	(0.402)	(0.464)	(0.465)
Entrepreneurial/Self-employed	0.947***	1.333***	2.090***
	(0.208)	(0.326)	(0.216)
Regional Corruption	0.053	-0.271	0.303*
	(0.147)	(0.273)	(0.177)
Interactions with Corruption			
Clerical × Corruption	0.253	0.608*	0.194
	(0.220)	(0.316)	(0.317)
Managerial × Corruption	0.983	1.174*	1.142*
	(0.625)	(0.686)	(0.662)
Entrepreneurial × Corruption	-0.101	-0.072	-0.010
	(0.221)	(0.404)	(0.202)
Education (years)	0.728***	-0.081	-0.146
	(0.183)	(0.317)	(0.169)
Education²	-0.013*	0.023*	0.011
	(0.007)	(0.012)	(0.007)
Observations	1,522		
Pseudo R^2	0.2303		
LR $chi^2(45)$	841.16***		

Notes: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ Reference category for the dependent variable: Manual occupations Fixed effects for macroarea-period and regional indicators included but coefficients not reported

Table A3: Sons vs Real Fathers Comparison (Restricted Sample)

	Sons	Real Fathers
Sample size	1,897	1,897
Age (mean)	41.923	41.923
Birth year (mean)	1965.448	1942.709
Education (%)		
None	0.26	12.30
Elementary	3.80	47.86
Lower secondary	36.90	24.70
Upper secondary	40.59	11.56
University degree	17.29	3.59
Post-graduate	1.16	-
Occupation (%)		
Manual worker	36.90	57.14
White-collar/Teacher	27.31	14.41
Middle manager	7.17	4.72
Senior manager	10.12	1.02
Professional	12.23	2.42
Entrepreneur	0.69	2.42
Self-employed/Family business	5.59	17.86
Income measures (mean)		
Log labor income	9.745	Not available
Log net disposable income excl. financial capital	9.956	Not available
Log total net disposable income	9.934	Not available

Notes: As described in the paper, fathers' income is not directly observed. The age of fathers corresponds to the same lifecycle point as their sons because respondents were asked to provide information about their fathers when they were the same age as the respondents are currently.

Table A4: Real Fathers vs Pseudo-Fathers Comparison

	Real Fathers	Pseudo-Fathers
Sample size	1,897	5,439
Age (mean)	41.923	43.018
Birth year (mean)	1942.709	1947.952
Education (%)		
No education	12.30	1.31
Elementary	47.86	22.32
Lower secondary	24.70	33.88
Upper secondary	11.56	31.20
University degree	3.59	10.74
Post-graduate	0	0.55
Occupation (%)		
Manual worker	57.14	33.64
White-collar/Teacher	14.41	30.46
Middle manager	4.72	8.23
Senior manager	1.02	3.42
Professional	2.42	4.68
Entrepreneur	2.42	2.08
Self-employed/Family business	17.86	17.50
Sector (%)		
Agriculture/Fishing	17.85	4.53
Industry	31.36	35.23
Public Administration	15.09	9.40
Other sectors	35.70	50.83
Geographic Distribution (%)		
North	40.29	38.59
Center	18.27	21.92
South and Islands	41.44	39.49
Income measures (mean)		
Log labor income	Not available	10.016
Log net disposable income excl. financial capital	Not available	10.217
Log total net disposable income	Not available	10.237

Notes: The real fathers lack income data. The sample size difference is due to the restriction applied to the sons sample to include only those whose fathers were born at least in 1939, making them comparable to the pseudo-fathers sample available.

Table A5: First-Stage Regression Results for TSTSLS Estimation

Variable	Net Total Income	Net Income excluding Financial Capital	Labor Income
Education Level			
Elementary	0.135** (0.046)	0.124** (0.045)	0.099 (0.057)
Lower Secondary	0.227*** (0.045)	0.224*** (0.045)	0.181** (0.057)
Upper Secondary	0.321*** (0.047)	0.316*** (0.046)	0.239*** (0.058)
Tertiary	0.507*** (0.051)	0.476*** (0.050)	0.352*** (0.060)
Post-graduate	0.682*** (0.103)	0.648*** (0.098)	0.506*** (0.102)
Sector			
Industry	0.133*** (0.034)	0.155*** (0.035)	0.285*** (0.038)
Public Administration	0.038 (0.037)	0.066* (0.038)	0.197*** (0.040)
Other Sectors	0.150*** (0.034)	0.181*** (0.035)	0.316*** (0.038)
Occupation			
Office Worker/Teacher	0.195*** (0.016)	0.179*** (0.015)	0.168*** (0.014)
Middle Manager	0.368*** (0.024)	0.338*** (0.023)	0.353*** (0.021)
Executive/University Prof.	0.657*** (0.037)	0.621*** (0.035)	0.643*** (0.033)
Professional	0.411*** (0.038)	0.379*** (0.036)	0.335*** (0.037)
Entrepreneur	0.530*** (0.058)	0.477*** (0.054)	0.367*** (0.052)
Self-employed/Family Business	0.227*** (0.021)	0.209*** (0.020)	0.109*** (0.020)
Geographic Location			
South (dummy)	-0.152*** (0.012)	-0.118*** (0.011)	-0.111*** (0.011)
Constant	9.747*** (0.050)	9.707*** (0.050)	9.441*** (0.064)
Observations	5,296	5,295	5,230
R-squared	0.298	0.294	0.281
F-statistic	156.50***	150.67***	131.04***

Variable	Net Total Income	Net Income excluding Financial Capital	Labor Income
λ for DC-TSTSLS	0.694	0.706	0.744

Notes: Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05. The dependent variables are log of net total income, log of net income excluding financial capital, and log of labor income, respectively. λ represents the parameter used for double correction in the TSTSLS estimation, calculated as $2(1-R^2)^3$.

Appendix B: Methodological Details

B1. Iterative Proportional Fitting

When comparing raw mobility rates (M) across different populations, observed differences may stem from structural changes in the occupational distribution rather than from genuine changes in the association between fathers' and sons' occupations. The IPF procedure addresses this limitation by standardizing the marginal distributions while preserving the intrinsic association patterns.

Consider a contingency table P with dimensions r x s, where rows represent sons' occupational categories and columns represent fathers' occupational categories.

The procedure operates through successive iterations that alternatively adjust row and column totals to match target marginals. For the k-th iteration, the adjustment proceeds as follows:

For row adjustment:

$$p_{ij}^{(k+0.5)} = p_{ij}^{(k)} x \frac{r_i}{\sum_j p_{ij}^{(k)}}$$

For column adjustment:

$$p_{ij}^{(k+1)} = p_{ij}^{(k+0.5)} x \frac{c_j}{\sum_i p_{ij}^{(k+0.5)}}$$

where $p_{ij}^{(k)}$ represents the value in cell $(i; j)$ at iteration k, r_i is the target row sum for row i, c_j is the target column sum for column j. In our analysis, the Center region was selected as the reference distribution, providing the target marginals for adjusting the North and South/Islands tables. We determined that $k = 10$ iterations provided sufficient convergence to the target marginals, with the sum of absolute differences between adjusted and target marginals falling below 0.1% of the total table frequency.

B.2 Derivation of the Relationship Between Altham Statistic and ANOVA

This appendix section presents the formal derivation establishing the equivalence between the Altham statistic and the Analysis of Variance (ANOVA) of logarithmic cell frequencies in contingency tables that has been exploited in this thesis.

Given the canonical log-linear representation of the contingency table $\log(p_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where γ_{ij} represents the interaction parameters capturing deviations from statistical independence, for any quadruplet of cells $(i; j)$, $(i; m)$, $(l; j)$ and $(l; m)$, the log odds ratio is expressed as:

$$\log\left(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\right) = [\mu + \alpha_i + \beta_j + \gamma_{ij}] + [\mu + \alpha_l + \beta_m + \gamma_{lm}] - [\mu + \alpha_i + \beta_m + \gamma_{im}] - [\mu + \alpha_l + \beta_j + \gamma_{lj}]$$

This simplifies to:

$$\log\left(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\right) = \gamma_{ij} + \gamma_{lm} - \gamma_{im} - \gamma_{lj}$$

Replacing this expression for the log odds in the Altham statistic, we obtain:

$$d(p, J)^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{l=1}^{r}\sum_{m=1}^{s}\left(\gamma_{ij} + \gamma_{lm} - \gamma_{im} - \gamma_{lj}\right)^2$$

Expanding the square and employing the identifying constraints of the log-linear model ($\sum_i \sum_j \gamma_{ij} = 0$, $\sum_j \gamma_{ij} = 0$, $\sum_i \gamma_{ij} = 0$), cross-product terms vanish, yielding: $d(P, J)^2 = 4 \text{ x } r \text{ x } s \text{ x } \sum_i \sum_j \gamma_{ij}^2$, where $\sum_i \sum_j \gamma_{ij}^2$ is just the residual sum of squares (RSS) from the log-linear regression.

B3. Projection Bias Correction

Starting from the variance-corrected estimator:

$$\text{plim}\ \hat{\beta}^{VC-TSTSLS} = \text{plim}\ \hat{\beta} - \eta, \text{ where } \eta \equiv \frac{\text{cov}\ (y^s, y^f - \hat{y}^f)}{\text{var}\ (y^f)}$$

Under the proportionality assumption:

$$\text{cov}\left(y^s, y^f - \hat{y}^f\right) = \lambda\ \text{cov}\ (y^s, \hat{y}^f).$$

After variance correction, $\text{var}(\hat{y}^f) = \text{var}(\ y^f)$:

$$\eta = \lambda\ \frac{\text{cov}\ (y^s, \hat{y}^f)}{\text{var}(\ y^f)} = \lambda\ \frac{\text{cov}\ (y^s, \hat{y}^{fv})}{\text{var}(\ y^{fv})}$$

Since the variance correction preserves the covariance structure:

$$\eta = \lambda\ \beta^{VC-TSTSLS}$$

Substituting back and solving for β:

$$\text{plim}\ \hat{\beta}^{VC-TSTSLS} = \beta - \lambda\ \beta^{VC-TSTSLS}$$

$$\beta = \hat{\beta}^{VC-TSTSTLS}(1 + \lambda)$$

Therefore, the double-corrected estimator is:

$$\hat{\beta}^{DC-TSTSLS} = \hat{\beta}^{VC-TSTSTLS}(1 + \lambda)$$

B4. Theoretical justification of the Lower Bound for λ

The projection bias parameter λ captures the relationship between unobserved factors in the first-stage regression and their intergenerational transmission. Since it is not directly observable, we need to derive a theoretically grounded approximation. The key insight is that the projection bias arises from omitted variables in the first-stage regression that are intergenerationally correlated.

The R^2 from the first-stage regression $R^2 = 1 - \frac{\mathrm{var}(v)}{\mathrm{var}(y^{pf})} = \frac{\mathrm{var}\left(\widehat{y^{pf}}\right)}{\mathrm{var}\,(y^{pf})}$ represents the proportion of variance in fathers' incomes explained by observable characteristics.

The relationship between λ and R^2 must satisfy several logical constraints:

Boundary Condition at Perfect Prediction: when $R^2 = 1$, all relevant factors are captured in the first stage, implying no projection bias. Therefore, λ = 0.

Monotonicity: as R^2 decreases (more variance remains unexplained), the potential for projection bias increases. Thus, $\frac{\partial \lambda}{\partial R^2} < 0$.

Convexity: the rate of increase in projection bias should accelerate as R^2 decreases, since each additional unexplained factor is more likely to be systematically important. This implies $\frac{\partial^2 \lambda}{\partial {R^2}^2} > 0$. Based on these constraints and empirical calibration studies, the lower bound is specified as: $\lambda_L = 2\,(1 - R^2)^3$.